AD-A170 941

Research Product 85-29

Comparison-Based Prediction of Cost and Effectiveness of Training Devices: A Guidebook

Training and Simulation Technical Area

Training Research Laboratory

July 1985

DTIC
ELECTE
AUG 1 3 1986
S D
D

DTIC FILE COPY

86 8 11 176

U. S. Army Research Institute for the Behavioral and Social Sciences

Approved for public release; distribution unlimited.

U. S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

WM. DARRYL HENDERSON
COL, IN
Commanding

Research accomplished under contract
for the Department of the Army

Klein Associates

Technical review by

D. MacPherson
S. Sabat

NOTICES

FINAL DISPOSITION: This Research Product may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: This Research Product is not to be construed as an official Department of the Army document in its present form.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER ARI Research Product 85-29	2. GOVT ACCESSION NO. AD-A170 941	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) COMPARISON-BASED PREDICTION OF COST AND EFFECTIVENESS OF TRAINING DEVICES: A GUIDEBOOK		5. TYPE OF REPORT & PERIOD COVERED Final Report April 1983-June 1985
		6. PERFORMING ORG. REPORT NUMBER --
7. AUTHOR(s) Klein, G. A., & Johns, P. (Klein Associates); & Perez, R., & Mirabella, A. (ARI)		8. CONTRACT OR GRANT NUMBER(s) MDA 903-83-C-0270
9. PERFORMING ORGANIZATION NAME AND ADDRESS Klein Associates 740 Wright Street Yellow Springs, OH 45387		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q162717A790 3413 102
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue, Alexandria, VA 22333-5600		12. REPORT DATE July 1985
		13. NUMBER OF PAGES 94
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) --		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE --

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

--

18. SUPPLEMENTARY NOTES

Contracting Officer's Representative, Ray Perez.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Maintenance training
Training devices
Training effectiveness assessment
Operator training
Transfer of training

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This Guidebook is part of a larger project to develop comparison-based prediction (CBP) techniques to estimate the effectiveness of new training devices and to generate design recommendations. The document explains why and how the methodology can be useful and then gives step-by-step instructions on how to apply it.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

A

Research Product 85-29

Comparison-Based Prediction of Cost and Effectiveness of Training Devices: A Guidebook

Gary A. Klein and Paula Johns
Klein Associates

Ray Perez and Angelo Mirabella
Army Research Institute

QUALITY INSPECTED 4

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

Training and Simulation Technical Area
Stanley Bolin, Chief

Training Research Laboratory
Harold F. O'Neil, Jr., Director

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

July 1985

Army Project Number
2Q162717A790

Training Devices and Simulation

Approved for public release; distribution unlimited.

ARI Research Reports and Technical Reports are intended for sponsors of R&D tasks and for other research and military agencies. Any findings ready for implementation at the time of publication are presented in the last part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

An ARI mission is to produce technology (i.e., job aids) that will help Army training developers design, acquire, and use simulation-based and computer-based programs of instruction for weapon operation and maintenance. A critically needed aid is one that will help the Army design and evaluate training devices early in the weapon acquisition cycle.

One approach to such aiding--comparison-based prediction (CBP) is the subject of this report. The approach has been used successfully as part of the HARDMAN method for estimating new hardware reliability. We are now trying to exploit CBP as a method for estimating the effectiveness of training devices as early as the drawing board or prototype stage of training development.

The current effort translates several years of scientific research and developmental testing into user-oriented guidelines. These guidelines can help PM-TRADE and training developers in TRADOC make better informed and documented decisions.

Edgar M. Johnson

EDGAR M. JOHNSON
Technical Director

COMPARISON-BASED PREDICTION OF COST AND EFFECTIVENESS OF TRAINING DEVICES: A GUIDEBOOK

CONTENTS

CBP GUIDE

CHAPTER I: INTRODUCING THIS GUIDE

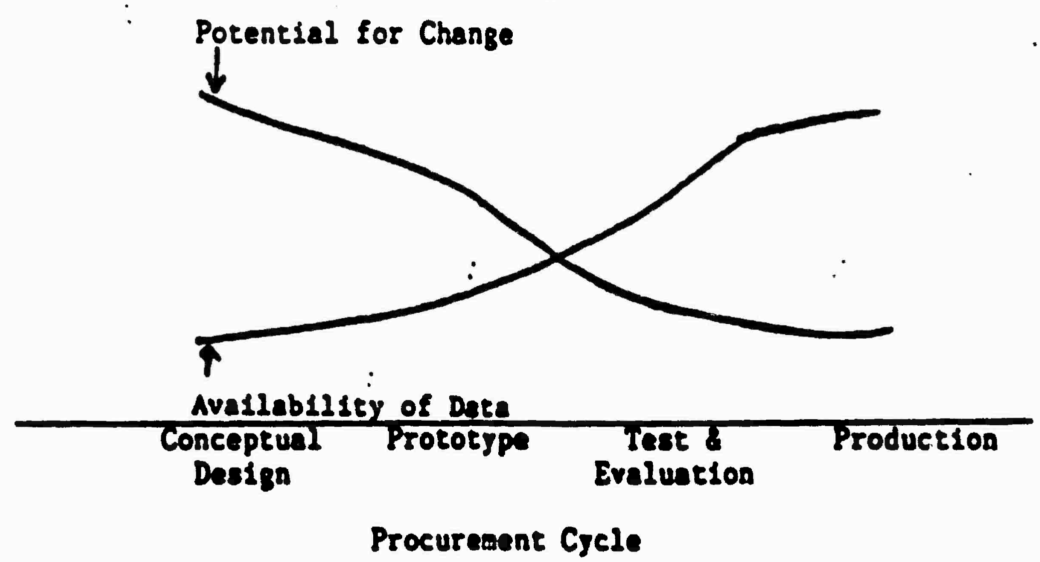
OBJECTIVE

The objective of this Guide is to explain to training device developers how to use the Comparison-Based Prediction (CBP) method to estimate the cost and effectiveness of new training devices. By training device we mean any equipment that represents or simulates the equipment and tasks of the system whose use is being trained. This can include the system equipment itself as an Actual Equipment Trainer. This Guide describes a way to make better predictions early in the design sequence.

Why do we need such a methodology? Consider Figure 1. It represents the fundamental challenge of the equipment procurement cycle: that the potential to make inexpensive changes declines as the procurement cycle progresses, while the data for deciding on these changes are not available until the cycle is almost completed.

Figure 1.

Potential for Change vs. Data Availability



Most cost effective decisions about the need for and design of a training device must be made early on. Decisions involving fidelity and function, type of training -- all have to be made almost before the system itself is developed, and well before the training device has been constructed and can yield operational data. Yet most prediction models require considerable data about the device, data which are not available until much later in the cycle. Therefore, a prediction method is needed that is not data driven, that can operate with information from sources other than the system under design.

Why is Comparison-Based Prediction useful? Because it uses experience with devices similar to the one under design, and therefore can be used early in the cycle, at the Training Development Study (TDS) or Cost and Training Effectiveness Analysis (CTEA) stages. It does not require operational data from the new system.

There are other techniques in use, ranging from reliance on expert judgment, at the simplest level, to mathematical models, at the most complex. But expert judgment by its nature is subjective, and it can be difficult to evaluate such predictions or justify them to others. And, at the other end, complex systems require large scale models to encompass all necessary factors that affect training effectiveness, and there may not be the operational data needed to feed such a model. While all the data and knowledge needed to feed prediction models may eventually become available, it is useful to have a reliable way to make decisions early in the cycle.

Comparison-Based Prediction provides a technique that is between unstructured expert judgments and complex data-driven models. It is a method for structuring expert opinion, so that it makes use of the data that is available about similar cases. It provides an audit trail of such judgments, so that they can be evaluated, compared with other predictions, and adjusted as the design process advances.

Who is likely to use CBP? Many people are involved in the development of training devices at different stages of the procurement cycle. Potential users may be personnel in Army training operations, responsible for the design and development of training programs or courses for new or modified systems. They may be personnel in contractor organizations, who must develop training requirements for new material. They could be personnel in program management offices (PMOs) responsible for training-program needs of new system development. Or they may be personnel in other military organizations, such as TRADOC system managers; or at other levels, such as TRASNA or PM TRADE, who have responsibility for training program development.

They need to answer such questions as: What type of training devices (for example, 3-D or 2-D?) should we consider for this system? How much should be budgeted for developing, operating, and maintaining a training device for this system? What will be the most cost effective training device for this system? What will be the characteristics of an effective device? Which training device will be most effective under these particular circumstances?

COMPARISON-BASED PREDICTION

What is this methodology? Comparison-Based Prediction (CBP) is a method of reasoning by analogy, where an inference is made for one object or event based upon a similar object or event. It is the use of concrete experience as a basis for predicting the future, making adjustments on the basis of key differences between the cases. Later in this manual, we will go into the method in considerable detail. And in Appendix A, we discuss the development of CBP and the research done with it to date. For now, let us describe this type of reasoning by a simplified, every-day example that will illustrate its principles and elements.

When homeowners decide to sell their house, they are going to do something they have never done before. They may have had experience in buying and selling other homes; however, selling their present home represents a unique, if similar, event. They need first to predict a realistic selling price — one which not only will attract a buyer but also will give them an optimum return on their investment.

They will turn to an expert, perhaps an appraiser, or a realtor who is knowledgeable about market prices for other houses bought and sold in their neighborhood. The expert will compare their house with others nearby that have sold recently or that are on the market, choosing ones that are like it in important features such as size, or age, or type of construction. The homeowners may decide to get opinions from more than one expert, or from persons with a variety of knowledge or experience. The general strategy is to seek expert opinion that is based on knowledge of comparable cases. The most important elements are identifying knowledgeable experts and finding appropriate cases to use for comparison.

The way an expert, such as a realtor or appraiser, proceeds is to identify the important factors that influence the selling price. These generally include those on which the selection of comparison cases was made, and other factors specific to the houses, such as number of bathrooms, or size of grounds, or orientation on the lot. Other factors in the situation are also noted, such as whether mortgage rates are rising, whether housing is scarce, whether the schools are good. The expert will finally winnow down a full list of important factors into those few, perhaps no more than five, that currently have the most influence on price.

The next step will be to assess the differences, between the comparison houses and the one under evaluation, on these important factors. The expert will examine them one by one, first to assess the direction of the difference, then to refine a measure of its magnitude. For example: the home may have four bedrooms, while the comparison case had only three. The extra bedroom will probably mean a higher selling price, but just how much is yet to be determined. At the same time, perhaps the home for sale has only one bathroom, whereas the comparison house had two. On this factor, the price may be expected to be lower than that for the comparison case. The expert later will fine tune the effect of these factors, with realistic adjustments based on experience.

The outcome is a price that started as the selling price on a comparable house or houses, adjusted up and down by an expert, on the basis of differences between this house and the other(s) on certain important factors. In addition, there is a clear record of this process. The expert did not pull a figure out of the air, and the homeowner knows the basis for the decision. The price can be justified to those who must evaluate it; and if there should be a change in one of these factors -- or if the experience with one of the comparison cases should turn out to be different -- an adjustment can easily be made.

This is an example of how reasoning by analogy works, and how we accept it as a useful methodology in areas that are important. Note that the expert could have developed a model, putting a price on each and every feature of the home and then adjusting for certain known factors. There are times when that is not practical or necessary; the use of analogy allows us to concentrate on only the most important variables.

While identifying many possible causal factors, the realtor settled on those few that currently were likely to have the most influence on price. Since a few factors generally account for the largest part of the differences between cases, the marginal value of additional factors is low. By using a well chosen analogy, we incorporate the vital elements of our subject, without having to know why they behave as they do, and focus on what is different about the new situation and what effect those differences will have. In effect, we have controlled for those many variables which are common to both cases.

Using the example of the home sale, let us define the elements of the CBP method. Then, we will build a model of the methodology. Table I lists the elements in the example of the home sale and shows the formal element of CBP methodology which each illustrates.

Table I

Elements of the CBP Methodology

Element in Illustration	CBP Element
The home being sold	Target Case: A
Selling price	Target Variable: T
Selling price for A	Target Value: T(A)
The realtor, perhaps other appraisal experts	Subject Matter Expert (SME)
Other homes, previously sold	Comparison case(s): B
Factors that may influence the selling price of A (e.g size, age, number of rooms)	Causal factors (from which high high drivers are selected)
Final list of most important factors, their specific values and how they affect one another	Scenario
Decision on how many comparison houses (B) to use and how many and what kinds of appraisors to use	Strategy
Selling price for a comparison house	Comparison Value: T(B)
Documentation/Report on how selling price of the target house was estimated	Audit trail

The steps in using Comparison-Based Prediction are summarized as follows:

PHASE I -- SET UP YOUR PROBLEM:

1. Specify the target, the device for which you are trying to predict cost or effectiveness. This is called A.

2. Define the measure of that cost or effectiveness. This is the target variable T; this is what you are trying to predict, what you need to know.

3. Identify the major causal factors (high drivers) that will affect the target variable for A, T(A).

4. Determine a context, or scenario, for your prediction, building in values for the high drivers. Under what conditions will A operate? How will the target value T(A) be measured?

PHASE II -- SELECT SPECIFIC RESOURCES:

5. Identify comparison devices B (1...n); if you are not knowledgeable here, you may have to consult with others who are.

6. Examine the CBP strategies to select the one most relevant to this problem.

7. Choose knowledgeable subject matter experts (SMEs), ones familiar with the comparison device if you have already chosen that. For some problems, you may need no expert but yourself.

PHASE III -- COLLECT YOUR DATA:

8. Determine, with your SME, the comparison value T(B). This will be the same variable, T, that you chose for the target device; but for the comparison case, the value, T(B), should be known already. In some cases, there will be no data available for T(B), and your SME will have to estimate them.

9. Present the high driver list. Examine the scenario differences between cases A and B. Estimate the effect of these differences, one by one, on the comparison value T(B).

10. Adjust the value of T(B) to allow for the differences between B and A.

PHASE IV -- MAKE THE PREDICTION:

11. Determine a value for T(A) from this adjustment.

12. Document this process to leave an audit trail. This becomes the basis for evaluating this decision, or revising it as development proceeds.

Why would you want to use this method? We have already discussed the need for predicting training device cost and effectiveness early in the design cycle. Here are some reasons for using this approach in particular:

ADVANTAGES OF CBP

CBP has characteristics that make it especially useful early in the training device development cycle.

DATA	--	It does not require a great deal of data about the targetfor which predictions are needed.
EXPERIENCE	--	The predictions are derived from operational experience, not theoretical models.
STRUCTURE	--	It uses expert judgments, but it structures those judgments to increase the quality of the predictions.
RELATIVE JUDGMENTS	--	It asks experts for relative judgments, that is, comparisons to other cases; these are easier to make than absolute predictions.
DOCUMENTATION	--	It creates an audit trail of the prediction process; the decision can be explained and justified and, as development proceeds, the prediction can be updated.

When is the use of the CBP method inappropriate? As any method does, CBP also has its drawbacks. These can show you when not to apply CBP.

DRAWBACKS OF CBP

DATA -- CBP requires data from specific cases, not summary statistics, and such case data may not exist. They may be estimated; and though subjective, they are systematically collected from experts who base them in their own experience, thus reducing error. If data for a formal model are available, then a reliable model should be used.

SUBJECTS -- CBP's simplistic model may invite too casual use. It requires expertise, but people disagree on who is expert. Using CBP successfully requires credible experts.

CASES -- Cases that are comparable to the one under consideration must be known for the method to be successful. However, it is rare that a device should be so unique as to have no similarities to others.

BACKGROUND

The Army has developmentally tested the use of CBP for predicting the effectiveness of training devices. Applications involved automotive maintenance trainers (AMTESS), videodisc gunnery simulators for tanks (VIGS), and trainers for self-propelled howitzer operations and maintenance (HIP). CBP methods have been used to predict such measures as time saved in training and effectiveness of training. A study at George Mason University, not yet reported, predicting the effectiveness of training devices with varying degrees of physical and functional fidelity, yielded correlations of .90 between CBP predictions and test results (Klein, in process). In another study, training personnel indicated more confidence in the predictions they made using CBP than in their own unstructured "expert" judgments.

Reasoning by analogy is already part of accepted military planning methodology. The Air Force has developed one procedure for such reasoning, Comparability Analysis (CA), to predict the reliability of new aircraft systems (Tetmeyer, 1976). Correlations between predicted and observed reliability of systems on A-10 aircraft ranged from .36 to .84, the higher for cases where empirical data were available and did not have to be estimated (Klein & Gordon, 1984).

To illustrate CBP in the Air Force, an engineer wishing to predict the reliability of the duct system of a new aircraft finds a duct system on a comparable aircraft that is already in use. The operational data on the reliability of the existing duct system serve as a data base. The engineer identifies differences between the new aircraft and the operational one that affect reliability. If, for example, the duct system of the new aircraft is twice the size of the duct system of the existing aircraft, and there are no other important differences, then the engineer may calculate that the data for the new system will be twice the magnitude of the existing data. This degree of adjustment of existing data is termed the adjustment factor, in this case 2X, or twice. The engineer applies this adjustment factor (that is, 2X) to the operational data to generate a prediction for the new duct system. Thus the prediction is based on the operational data but enhanced by the engineer's judgment of how to adjust those data to fit the new situation (in this case, to multiple them by 2).

For a more comprehensive discussion of the development of CBP, please see Appendix A.

This has been the who, what, why, and when of CBP methodology for predicting training device effectiveness and cost. We have described a process that structures expert judgment, using data from comparison cases from the expert's own experience, and providing an audit trail for evaluation and update. The method is suitable for use early in the design stage, when there is insufficient data for the use of formal prediction models. We are now ready to turn to the how of this process, and detail the steps in the CBP procedures.

CHAPTER II: HOW CBP WORKS: A Quick Example

Back on page 1-8 we summarized the steps of CBP. We have turned that summary into a flow chart, on page 2-2. Both of these summaries can be referenced as you follow the hypothetical example below of a CBP application. In addition, Appendix B contains sample forms that support this example.

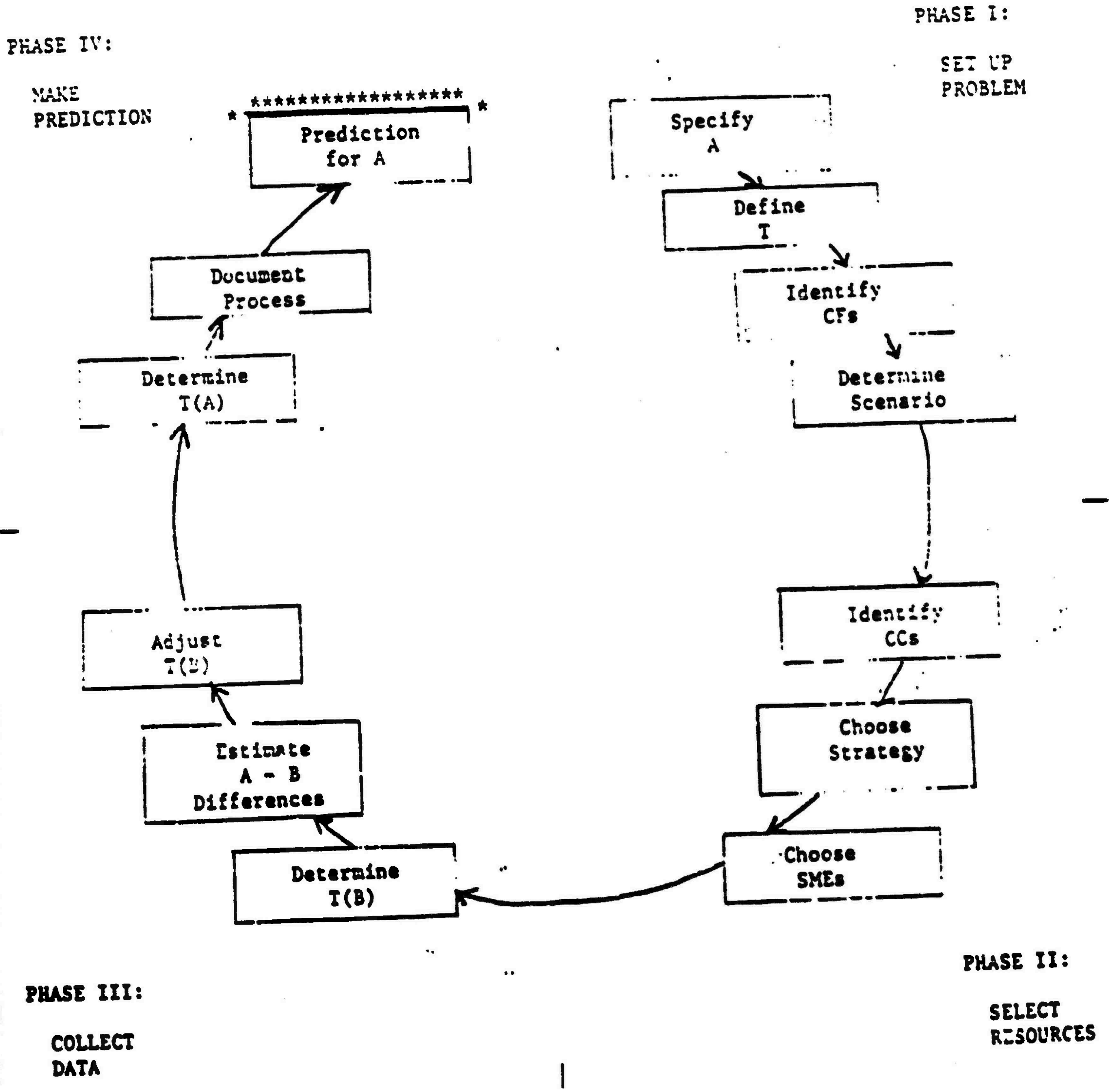
A word about the diagram on page 2-2. The flow chart is circular, suggesting the iterative nature of the process. Sometimes information received in one step will prompt you to return to an earlier step to rethink or adjust a decision made there. For example, you may choose a comparison case and, in a subsequent step, the expert you choose suggests a better case that you did not know about. These steps are not to be taken as rigidly sequential. Also, the same resources may be used at different stages of the process.

This circular flow chart of the process is also a visual aid to the organization of the remainder of the Guide. You will see it printed at the top of each section, with the shaded portion of the diagram being the steps under discussion in the following text. It will be easier to find your way in the process by making reference to the diagram on the page you are studying.

Throughout the example application in this chapter are inserted lists identifying the example's elements in terms of the elements of CBP methodology. To assist your understanding, we also repeat the identification of elements from the home sale example. And for each element, we also show its alpha symbol.

Figure 2

Comparison-Based Prediction Process



Copy available to DTIC does not permit fully legible reproduction

SETTING UP THE PROBLEM

The Army is building a new howitzer, and you are charged with recommending a training device to be ready for use when the howitzer is delivered. There are many decisions to be made: whether to use an existing device or build a new one; whether to use a simulator or recommend training on the actual equipment; and many others that would come before and after this level of recommendation. Assume that for our example you have narrowed in on trying to predict the effectiveness of a 3-dimensional, highly realistic training device that will have many instructional features, such as feedback mechanisms and recording devices. You have decided that the measure of training effectiveness will be the number of direct hits the class members average on their final test round.

CBP ELEMENT	SYMBOL	IN THIS EXAMPLE	IN HOME EXAMPLE
Target Case	A	Training device for the new howitzer	Home being sold
Causal Factors	CF	Physical Fidelity; Feedback Potential	Size, age, etc.
Target Value	T(A)	Average hits, final test round	Selling price for A

SELECTING SPECIFIC RESOURCES

You investigate and find that there is currently no device with this degree of fidelity for training on the howitzer. You are not sure whether to choose as a comparison case a howitzer training device that lacks physical fidelity, or a training device for a tank gun that is close to your subject in that it has the important features of physical fidelity and feedback capability. You choose the tank gun training device but decide to use also a howitzer panel trainer that has good feedback characteristics.

CBP ELEMENT	SYMBOL	IN THIS EXAMPLE	IN HOME EXAMPLE
Comparison Case(s)	B	B1 Tank gun simulator B2 Howitzer panel trainer	Other homes, sold or on the market

You decide to call on training supervisors as your SMEs, and choose two for their respective experience in using the B1 and B2 training devices.

CBP ELEMENT	SYMBOL	IN THIS EXAMPLE	IN HOME EXAMPLE
Subject Matter Experts	SMEs	#1 Tank gunnery training supervisor #2 Howitzer gunnery training supervisor	Realtor, appraiser

You interview each SME separately, and have the SME consider only the training device which is the most familiar. [NOTE: this is a choice of strategy.]

COLLECTING THE DATA

Before starting the interview you draft a guide for it, so that you will present each SME with the same description of your proposed training device and give the same definition of the causal factors (e.g. physical fidelity, instructional features) you have identified. You also want to list the questions you will ask to be sure you cover the same ground with each one. You can plan on an interview of 45-60 minutes.

You find that in the training courses for howitzers and for tank gunners, there are different designations for what you thought of as an appropriate measure: number of hits on final test round. Each SME led you through the intricacies of training and testing, and you wound up with considerable refinement in your description of the comparison target variable and the circumstances under which it was measured.

CBP ELEMENT	SYMBOL	IN THIS EXAMPLE	IN HOME EXAMPLE
Comparison Value	T(B)	You have specified for each of B1 & B2 the level of class, specific round, and other details that would equate T(B) as nearly as possible to T(A) as a measure of each training device's training effectiveness	Selling prices for the comparison cases

You and the SME review possible sources for finding operational data for the performance of the level of class you have chosen. The Army, you discover, does not collect or keep these scores in any systematic way. SME #1, the tank gunnery training supervisor, happens to keep these data informally, as a way of checking out his instructors, so you can calculate a value for T(B1). SME #2 does not keep these data for his howitzer training courses; however, he feels pretty confident about estimating them for you, because he goes over the scores before winding up each course and has a good feel for them.

CBP ELEMENT	SYMBOL	IN THIS EXAMPLE	IN HOME EXAMPLE
Comparison Target Values	T(B1)	14/20 = 70% final test round hits	Selling prices for comparison cases
	T(B2)	estimated 75% final test round hits	

You then lead each SME through an examination of the differences between his training device and the one you are considering. You describe yours, and stress the high drivers you have identified: physical fidelity, feedback provisions, performance data recording.

SME #1 feels that, based on those factors, your training device should train about as well as his does, and he would predict the same 70% performance for your classes. But he thinks of something else: the computer that his feedback and recording features operate from has always been poor, and is down so often that he feels it has impeded training. You both agree that by the time your training device is operational, computer support should be much more dependable. He thinks this would produce better training and raise the average class score by a good 5%.

SME #2, in his interview, also examines the two training devices and the causal factors, including the new one that SME #1 identified. His training device is newer than SME #1's, and he is satisfied with computer dependability. He doesn't think that full physical fidelity is going to add much to training effectiveness, because he says his classes get in some practice on actual equipment before they take their final test rounds. But he thinks the performance recording system that you are building in will be not only a waste of time but an actual deterrent to training. He thinks you are going to waste time, load your instructors down with unnecessary tasks, and in general distract from what he considers to be the heart of effective training: drill and practice. He therefore thinks that, all things considered, your device will be slightly less effective than his is, maybe 5% so.

MAKING THE PREDICTION

You now can make a prediction of your target value: the number of hits you expect the class, using your planned new training device, to average on their final test round. The first SME thought it would be another 5% higher than his own classes' 70% average: that is, he thinks instead of 14/20 hits, they would have 14.7/20 hits, or:

CBP ELEMENT	SYMBOL	IN THIS EXAMPLE	IN HOME EXAMPLE
Target Value	T(A)	estim'd 73.5% final test round hits	selling price for home being sold

SME #2 had estimated his class at 75% average scores. He thinks yours will be 5% lower, or:

CBP ELEMENT	SYMBOL	IN THIS EXAMPLE	IN HOME EXAMPLE
Target Value	T(A)	estim'd 71.2% final test round hits	selling price for home being sold

You now have two expert opinions of what the training effectiveness, expressed in a class performance score, of your proposed device will be. The two scores are close, and you could simply average them and use 72.4% for a working figure.

However, you have documented these interviews. (Remember the guide you drafted before the interviews? You noted the answers on it.) Therefore, you are able to refer to your notes and try to resolve any discrepancies in judgments. And if your review supports feelings you may have had that you were not getting really responsive answers, you are free to reject an interview altogether.

In this example you tend to think, looking over their responses, that you trust SME #1 a little more, because he keeps records. So you think, if anything, your estimate may be a little low because of that. Whenever you use the 72.4%, you will keep that in mind.

Your working figure for this prediction is about 73%, with an expected range from 71% to 75%. The two different estimates become like confidence limits for the prediction.

CBP ELEMENT	SYMBOL	IN THIS EXAMPLE	IN HOME EXAMPLE
Target Value	T(A)	73% +/- 2%	selling price for home to be sold

CHAPTER THREE: SETTING UP THE PROBLEM

SELECTING THE PREDICTION TARGET

It is not always easy to decide just what you wish to predict. When you are attempting a cost prediction, it is usually not difficult to specify your variable. But if you want to know the effectiveness of a prototype maintenance training device, you must decide just what measure to use and what specific data you need.

The prediction variable, T, must be defined in terms of a clear question. If you want to predict training effectiveness, you might begin by asking yourself, "If the training device does a good job, what measure will show a big before/after training difference?" For example, if you mean effectiveness for training in a classroom, the T(A) might be the number of hours needed to cover certain tasks. Or, it might be the accuracy or speed of student performance on a particular task, or some other measure. The important point is that you want to be specific about what you mean by training effectiveness for a given case, and you must specify the measure of it.

You must also think about who will be using your prediction, and how it will be applied. This will give you ideas about how precise your measure must be, and what form it might need to take. Do you want to compare this device with others? Or do you want to use this device in the solution of a particular training problem? Table II is a checklist of items to consider in selecting T. You may add to it as you proceed with your specific problem. Table III lists some general measures for relative comparisons.

TABLE II

Checklist for T(A) Selection

__What do you need to predict?

__Are there standard measures?

__Does T(A) need more than one measure?

__Who will use this result, and how?

__Does your measure reflect training device use?

__How will you obtain T(B) data?

TABLE III

MEASURES OF COST	MEASURES OF TRAINING EFFECTIVENESS
investment costs	accuracy
operations and support costs	recall
instructor	speed of performance
facilities	transfer of training
maintenance	savings
life cycle costs	recognition
	performance on secondary task
	effort/efficiency
	number of wins (gaming task)
	number of instructors needed
	amount of supplementary actual equipment training needed
	skill decay curve
	time to criterion

One factor that can guide the selection of T is the availability of comparable data for T(B). That is, when you choose the target measure for predicting about A, you are going to have to find the value of that same measure for your comparison case B. Your knowledge of available data for B may influence your choice of measure. In predicting costs, for example, you know that there are cost data on record for comparison devices. When you want to predict training effectiveness, you may well have a problem finding comparable

The DoD does not maintain easily accessible records of the training effectiveness of existing devices, although sources are available. It is worth thinking about data availability before you decide on T. Sometimes you can find empirical studies of the effectiveness of comparison devices, and you can use the data as T(B). More often there will be no such studies. At this point you may bring in a subject matter expert for help (not necessarily the same one who will work on the actual prediction).

The list of possible comparison cases (Bs) does not have to be shortened until later in the CBP process; the final choice is not firmly made until the interview, when that SME agrees to it. However, it is useful to take T(B) into consideration early in your thinking. Knowing that you may have to estimate data for the comparison case will help you define your target variable: that is, it will encourage you to define a measure about which comparison data can be estimated, and it will start you thinking early about how to get those estimates.

You will find at times that you have to develop a very detailed measure, and that this will require creating a whole scenario of events within which to estimate comparison case data for it. Here is an example of where the choice of T became very complicated:

In predicting the effectiveness of a tank gunnery training device (the Videodisc Gunnery Simulator or VIGS), we had to construct a measure of skill sustainment, and came up with the following in consultation with our SMEs. Normally, there is no practice in the six months between the end of training and the start of field exercises. We wrote a scenario in which the men were to practice on the VIGS during these six months. We defined the target value T(A) as the number of first-round hits a trainee who had practiced with VIGS might be expected to achieve at the start of field exercises. For a comparable T(B), our SMEs had to estimate the number of first-round hits trainees would average if they had practiced with another training device.

There were no data, of course, for T(B), nor even for first-round hits attainable after six months with <u>no</u> intervening training. So SMEs first had to estimate the number of first-round hits the men would have achieved at the end of training, and then estimate the decay of an intervening six-months without practice, in order to estimate the effect of practice with other training devices.

LISTING THE CAUSAL FACTORS (HIGH DRIVERS)

All those characteristics of the device or its context that, differing in value from one scenario to another, can account for differences in the measure of T, are called causal factors. Those few factors, perhaps no more than five or seven, that account for the majority of the difference, can be termed high drivers.

No comparison case will be a perfect match for the training device you are planning. The two training devices will certainly differ if you are planning one that incorporates new features, or leading edge technology. Therefore you must give some thought to the most influential factors that could affect training. Central to the CBP methodology is the SMEs' ability to assess the impact of differences in these factors, especially to judge the impact on the variable of interest, T.

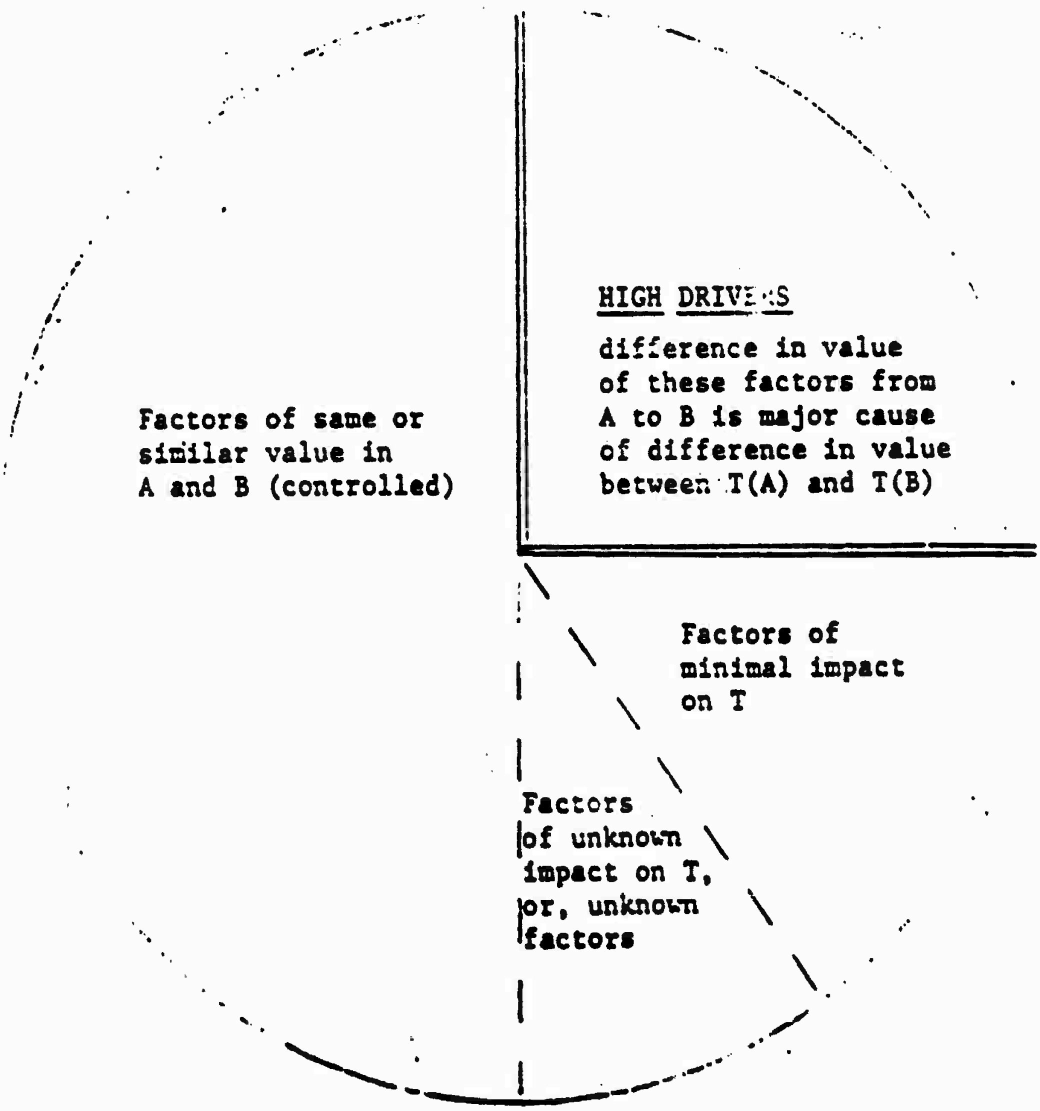
For example, if you are under pressing time limitations, the best first estimate of T(A) may be simply the value of T(B). Many ball park estimates of cost are made by copying cost figures from earlier programs. Such a figure is probably a better prediction than one based on an open-ended unstructured estimate. But CBP improves the prediction by utilizing experience with training devices and structuring the experts' ability to judge the effects of differences between the new situation and the one they have experienced.

Figure 3 illustrates why we need attend only to the high drivers in our comparison of the differences between cases. For many, perhaps half, of the characteristics we identify as causal factors, the impact of the scenario will be the same on the comparison case factors as on the new devices. For example, the physical durability of a simulator

Figure 3

CAUSAL FACTORS

Examine only the High Drivers in Comparison Based Prediction



can be relevant to its effectiveness, but in the absence of any information we assume that it will not differ from case to case and so do not create a scenario value for it. Physical durability is a given and thus of no interest in our analysis.

For a smaller percent of the factors involved, the impact of differences in value is minimal and so not worth our effort to calculate. Another small portion of the factors are likely to be unknown to us and so cannot be examined in any way. These unidentified factors may be the source of error in our predictions.

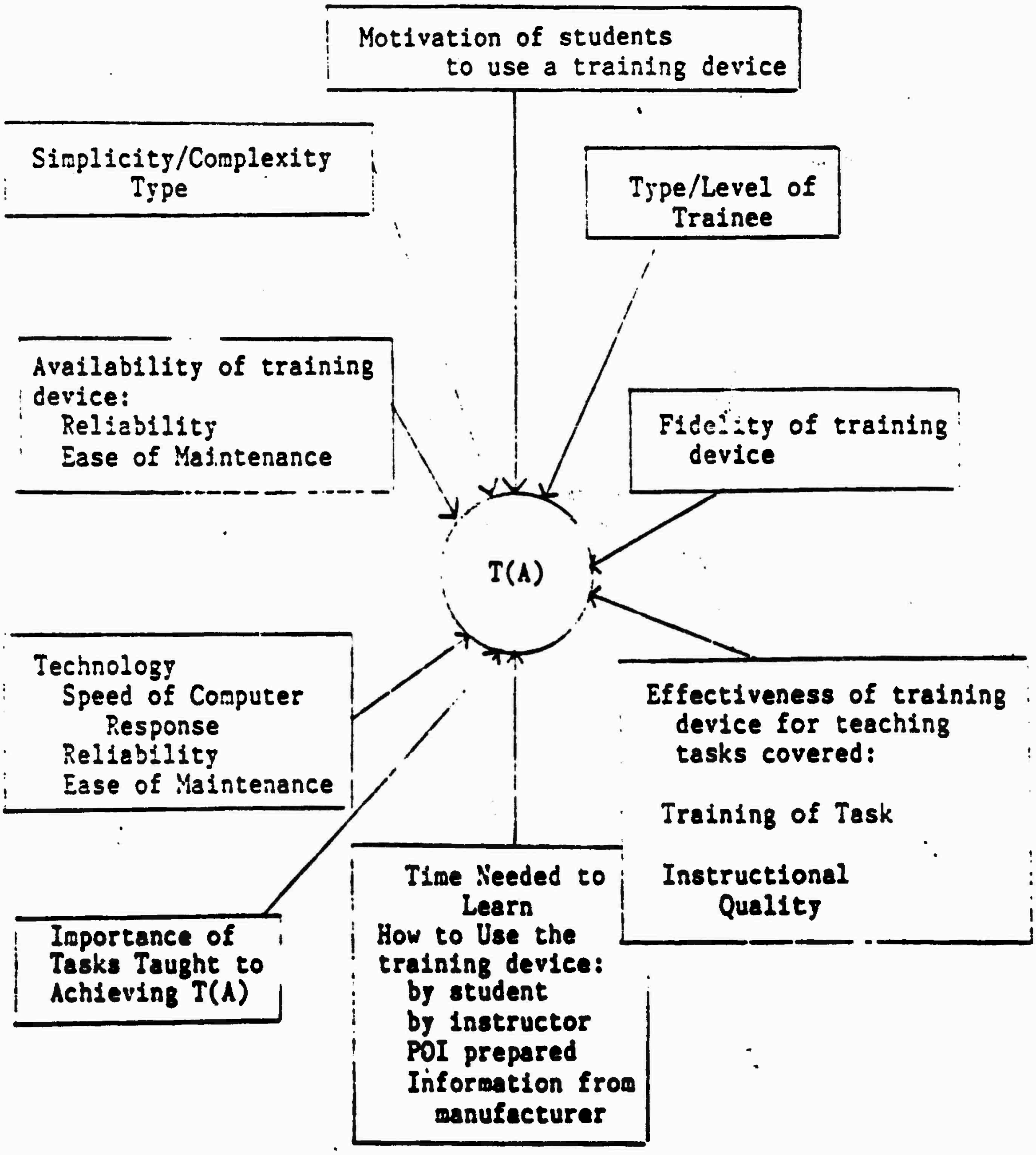
Thus only a selection of the causal factors will qualify for our attention as high drivers: those which, differing in value from case B to case A, will account for the major differences between T(B) and T(A), as Figure 3 illustrates.

The data we have found or estimated for T(B) can be thought of as experience. The SME will modify that experience to predict to a new case, T(A). The identification of the causal factors is the first step in transforming our experience with B into a prediction for A. The way that modification of experience is done depends on the causal factors and how they differ in value in the scenario for A.

You may need to consult an SME for this step, who may be someone other than the SME in your interview. Create a checklist of the most important causal factors that will affect T(A). Then examine these to see which ones were considerably different in case B. Remember, also, that we are looking at these factors in a specific context. While you will probably begin with differences between the two training devices, there may well be differences in the context — such as level of student to be trained — that will have sizeable effect. Figure 4 illustrates this concept.

Figure 4

ILLUSTRATION: SOME CAUSAL FACTORS AFFECTING
THE MEASURE (T) OF TRAINING DEVICE EFFECTIVENESS



Some differences in equipment will have significant impact on your target variable. For example, the addition of a feedback mechanism, that allows the trainee to know performance accuracy and to identify errors, can be expected to affect the learning process significantly. Other features of the training environment, influenced by the training device design or simply part of its operations, might include: student motivation; the simplicity or complexity of the task; availability of the training device; the level of student; instructional quality; or the realism of task presentation.

Table IV is a general checklist for your reference.

Try to limit the list of high drivers to five or seven, in order to simplify the job of the SME in the interview. But if in doubt, leave the inclusion of a factor to the SME's judgment. And try to define the high drivers so everyone will know what they mean, so that the distinctions between them are clear.

An example of the listing of high drivers, and the nature of their possible effects, comes from a study predicting the effectiveness of automotive repair training devices. The proposed training device was to be for the Army, but the comparison case was an Air Force device. The high drivers that were identified included the following:

#1 Computer Power	speed should increase effectiveness by both increasing motivation and providing increased training time
#2 Task Simplicity	simplicity should increase the chances of building an effective training device
#3 Trainee Level	less experienced students are less comfortable with simulators, more trainable on actual equipment

Table IV

ILLUSTRATION: SOME CAUSAL FACTORS AFFECTING THE MEASURE (T) OF TRAINING DEVICE EFFECTIVENESS
(Taken from a Tank Gunnery Training Device Application of CBP)

Causal Factors

Physical Fidelity
- Visual
- Audio
- Feel of controls

Device Utilization
- General ease of use
- Set up ease
- Equipment characteristics
- Frequency
- Computer Power

Training Value
- Instructional aids (kill zone)
- Feedback
- Familiarity/confidence
- Relaxation
- Motivation

Functional Fidelity
- Gunner training
- Procedures - automatic
- Range of tasks
- Variety and quality of training problems
- Lead training
- Malfunctions
- Trainee Level
- Task Simplicity

Task Characteristics
- Task Simplicity
- Crew Coordination

Trainee Level

Miscellaneous
- Cost
- Range

These three factors were judged the most powerful differences between the prediction case (Army) and the comparison case (Air Force). They illustrate three kinds of high drivers (in accord with the taxonomy in Table IV: #1 is a training device equipment characteristic; #2 is a task characteristic; #3 is a trainee characteristic.

To complete this example, although getting ahead of this chapter, we should tell you how the SME judged the effect of each of these causal factors. T was a measure of training time, expressed as class hours. Because the Army computer was to be faster, the same course could be taught in less time, thus reducing T(B). Because the task that the Army would use the training device for was to be simpler, the time to teach would again be reduced. However, the Army was going to train students less experienced than the Air Force had, so the SME added some to T(B).

This is shown in the following results:

	T(B) Air Force	T(A) Army
#1 Computer Power	-	+ (10% less time)
#2 Task Simplicity		+ (16% less time)
#3 Trainee Level	+ (13.5% less time)	
Overall		+(12.5% less time than Air Force)

The outcome was a prediction of a 12.5% savings in training time for the projected Army training device, compared to training time for the Air Force training device. Since we knew the length of time the Air Force training took, we could predict a time savings for the Army.

CONSTRUCTING THE SCENARIO

Prediction must be made about events in some context. A training device functions in a particular context with a defined range of task demands. In order to predict how well a training device will work, or what the value of a target variable will be, the person making the prediction must know how the device will be used or under what conditions the measure will be taken. (MIL-T-29053B guides you in describing instructional system development activities, including a functional description of a training device.) It is a formal element of the CBP method called the scenario, or frame.

If you do not establish these details about the device A — how it will be used, what tasks will be trained with it. etc., — variations can occur from one interview to another. Worse, the comparison case B may not be truly comparable with respect to important details, such as type of trainees or conditions of training. You would then produce 'findings' that would be misleading or have no real value. For example, specify if it will be used to train new inductees, or to introduce officers to a new specialty.

The scenario at a minimum should include the high drivers that you have selected for this situation. It is in the scenario development that you assign values to them. For example, every device will have trainees; the level of trainee is specified in the scenario.

Table V lists some of the factors to consider in developing this scenario. Some of the factors that our experience has shown to be especially relevant are a statement of the level of physical fidelity of the training device, of the convenience of using it in a class, and of how easily the training device might be modified to accommodate teaching different tasks as new ones are added to the course.

The scenario must provide a specification of the conditions under which the training device will be used, so that one or more experts (SMES) can judge its effectiveness in that context relative to other training devices. If the device may be used in different circumstances, which will have significantly different effect on its usefulness, the scenario may be appropriately modified for a review of those differences.

TABLE V

(SOME) FACTORS TO CONSIDER IN DEVELOPING THE SCENARIO:

TRAINEES	__Who are they? __What is their experience level? __Are they similar or dissimilar to current trainees?
TASK	__What is to be trained? __Is the task specific? or generic? __What are the criteria for learning?
PROGRAM	__What equipment is involved? __Where is it administered? __Are there options in the unit? __If in a school, what is the course? __How will it be modified to include the training device?
DEVICE	__What physical features does the training device have? __What instructional features? __What level of descriptive detail is needed?

CHAPTER IV: SELECTING THE RESOURCES

CHOOSING THE COMPARISON CASE

In this stage of the CBP process you will be making a tentative selection of one or more potential comparison cases (B) and choosing the Subject Matter Expert (SME) whom you will interview to obtain the needed information. These procedures are closely interrelated and must be pursued more or less together.

The first step is usually to prepare a list or menu of comparison cases. What this requires is that you identify other training devices, including Actual Equipment used as Trainers (AET), and gather information about each one -- where it is used, what tasks it trains, for what level of trainees it is designed.

It will likely be necessary to go outside your own experience to identify training devices that have similarities of form or function to the device which you are considering. You should not limit yourself unnecessarily in this task, but should consider tasks or equipment that are similar to your own system, and consider devices that may be in use in other commands and even other services.

Again, you may want to call on a consultant SME for this task. Remember, you need not be expert about the device you eventually will select. This is an opportunity to seek information about the full range of training devices that may be in use.

To select a good comparison case, use the list of high drivers to match it to your target case. The two should be as similar as possible on these important causal factors. Physical fidelity and task type are examples of high drivers that should be closely matched for: this is because their effects are so complex and important that the SMEs could not realistically assess the effects of differences in them.

As you identify possible training devices, maintain a data sheet on which you enter the characteristics of each, so as to facilitate the comparisons and the choice of a final subject. (Samples of such a form are included in Appendices B and C). Be sure to include those characteristics that you have already identified as likely causal factors. As you identify other elements of new training devices that you learn about, you may get further insight into what the high drivers should be.

While the job of identifying potential comparison training devices is yours, the final choice must be made in conjunction with your SME — who, by definition, is the expert on the topic. However, the final choice of SME depends in part on your choice of comparison training devices, because you are seeking expertise on the training device, not expertise on a general subject. It is possible that, having identified likely training devices and chosen an SME who is knowledgeable, the SME may then suggest yet another training device for use as a comparison case. It may be an entirely new training device, one that you had not discovered. Or it may be one you had considered, which the SME knows well enough to value more highly than you did for use as a comparison case.

This is a situation where you want to use the expertise of the SME, once he or she understands the problem. Guide the SME to select: 1) a case with which he or she is familiar; 2) one which will be a good match for the target case with respect to the high drivers; 3) a case for which data are readily attainable. This task seems a circular exercise. Our circular system diagram, Figure 2, illustrates that it is!

SELECTING THE EXPERT

Since Comparison-Based Prediction relies on subjective judgments, you need to find reliable and experienced personnel to make these judgments. All prediction methods rely on SMEs at some point, either for the development of a model or for the estimates that go into it. This is a critical element of the process. It is essential to select individuals who will be able to make accurate predictions and also in whose judgments others will have confidence. Without credibility, predictive accuracy is useless.

To some degree, the choice of SME will also depend on the way you have defined your variable T. In the last analysis, the job for the SME is to adjust the data for T(B) so as to produce a predicted value for T(A). You may elect to use training personnel who conduct the courses in which the comparison training device is used, and a training person may indeed be expert in its practical aspects. An SME must, at a minimum, have the ability to conceptualize T(B) as you have defined it. If you have devised some hypothetical measure of effectiveness for the SME to estimate, you must have an SME who will understand the objective you are aiming for and be able to abstract his or her experience for your needs.

If you have located research data on effectiveness, you may want to use research professionals as SMEs to adjust it. If you are working with operational data and estimates, you will want to use people with direct training experience, who will understand the context in which the data were derived and the variables that will be important as you pose changes to that context. Some sources for SMEs include ORSA, DOTD, school instructors, ARI research personnel, NTEC, etc.

If you have enough time and resources, you can use several SMEs: training researchers; systems designers; course instructors and supervisors. We will discuss this further when we consider different strategies for CBP.

ALTERNATIVE STRATEGIES

There are a number of ways in which the CBP application can be structured. Your choice of strategy will depend on such factors as time constraints, abundance of comparison cases, availability of data, and identification of SMEs.

Global Strategy. In the simplest application of the method, you will interview one SME, present all the relevant data about A, including the list of high drivers, and ask the SME to make a judgment about T(A) by adjusting what he or she knows about T(B). You may try to elicit the answer in terms of an actual value for T(A), or you may receive only a statement about an adjustment to T(B). But no further analysis would be undertaken. This is called the global strategy.

> As an example, you may be concerned to predict training hours needed for a specific task. The SME knows that a comparable training device uses 20 hours. Examining the high drivers and the description of A, the SME may judge "25 hours for A" or may go so far as to say only "maybe a fourth again the time for A." In this simple strategy, you would ask the SME only for his rationale; which one or two high drivers he considered.
>
> In another example of the global strategy, estimating the cost of a Howitzer Fire Control Panel Trainer, the SME selected a Multiple Launch Rocket System Trainer as the comparison case B and used its cost as T(B). She decided that for purposes of obtaining a ballpark figure, the effect of the high drivers could be discounted, and plugged T(B) directly in as T(A).

High Driver Strategy. In a more in-depth version of the CBP interview -- the high driver strategy -- the SME is asked to detail how target case A and comparison case B differ from each other. The SME is given the checklist of high drivers and judges whether each of these differences favors the target case A or the comparison case B, and whether any offset each other. The SME next estimates how much each of the high drivers affects the difference. The sum of these estimates is calculated during the interview. The SME then checks this calculation and is given opportunity to adjust it for reasons that must also be specified.

> An example of this strategy would be found in Chapter 2's example of the howitzer prediction, if we had kept to only one SME. T(B) was adjusted for three specific high drivers, one by one.

Multiple Comparison Strategy. Several comparison cases are used in the multiple comparison strategy. Basically you play a game of (fewer than) twenty questions to pin down the value of T(A) by progressively narrowing the prediction range. If the value of T(A) is smaller than T(B) for example, the prediction may be further clarified and refined by comparison to other cases also judged to be smaller than T(B). You need only two or three comparison cases before you have the prediction fairly well determined.

The use of multiple comparisons can serve to clarify and refine the judgment about the target measure. It can also serve to increase confidence about that judgment: if the same prediction is reached independently through the use of different comparison cases, we feel more certain about its accuracy. Also, if we use the judgment of multiple SMEs and their judgments converge on the same prediction, our confidence is raised.

As an example of the use of both multiple comparison cases and multiple SMEs, we have the AMTESS study applying CBP to predicting effectiveness of automotive maintenance trainers in specific courses. There we used two sets of SMES: Army instructors who were trainers with the actual equipment, and Air Force professionals familiar with comparable training devices. One case yielded a prediction of a 17% savings in course time. The second yielded a prediction of 13% savings in course time. These predictions were accepted as in the same ballpark, and a composite prediction of 15% savings, +/- 2%, was used to describe the results. The composite prediction was considered better than the single prediction because it included two different sources of SME input yielding results which converged at a reasonably close value for T(A).

This example, combining two strategies, can be called a convergence strategy.

By having more than one comparison case you can usually build into your prediction more causal factors for analysis -- that is, new cases will probably have features that are different from other cases, and so you may be able to find direct comparisons on more of the features of your proposed TD. Remember, you are trying to locate experience comparable to your prediction case: the more experience with more variables that you can find, the better your basis for prediction.

When using multiple comparisons, try to keep the interviews from becoming too complex. SMEs should be asked to make predictions based only on the training device with which they are most familiar. If they are really experienced with more than one, then be sure you reduce the list of causal factors to a number that can be handled without confusion. Be very clear in the interview just which comparison case you are talking about at any time, and that you are comparing it to the proposed training device.

Table VI lists some strategy options.

Table VI

Strategy Options

STRATEGY	ADVANTAGES	DISADVANTAGES
Global	Low resource demands	Weaker prediction Less explicit audit trail
High Driver	Explicit audit trail Causal factors evident	Difficult to use with several SMEs and/or multiple comparisons
Multiple Comparison Cases	Structured predictions	Requires more time, and availability of multiple comparison cases
Multiple SMEs	Broader input, cross-checks possible	Requires more time and SMEs Resultant prediction is complex

Cumulative Strategy. You may also elect a cumulative strategy, adding SMEs until you have enough agreement to feel confident. If you collect data from more than one SME, you can cross-check their results to see if there are disagreements. (NOTE that you should run only one SME at a time. When two or more are interviewed in the same session, the judgments of one are often dominant.) By examining the audit trail, you may find the basis for the disagreement in a misunderstanding, or in a difference of perspective, or in additional information that one had and the others did not.

If you can identify a reason for the differences between SME judgments, you can feel more confident in your ability to evaluate their combined judgment -- that is, you can decide whether it is wiser to take it on the high side or low side, etc. And you have more confidence in the judgment because you can evaluate the degree to which the experts agree with each other, knowing on what the differences in their calculations were based.

If, however, there is one SME whose judgment you value most highly, then you should use only this SME. Collecting more data will only add confusion. There is bound to be some difference between SMEs. Unless you know you are going to add value by having more than one opinion, you should not seek additional ones.

If there is time pressure for the prediction, and limited resources available for the CBP process, then select just the one best SME.

CHAPTER V: COLLECTION AND ANALYSIS OF DATA

Now that you have selected your SME and decided on the strategy you will use to obtain the prediction you need, you are ready for the data collection interview. This should take place face-to face — although it can, if necessary, be conducted over the telephone to save resources. Be extra careful about consistency here: you lose the chance to catch facial expressions that may show the SME's doubt or confusion about a step or question.

PREPARING FOR THE INTERVIEW

As you follow the CBP method and conduct interviews with SMEs, you will need to refer to different kinds of information. Prepare your materials well in advance and have them on hand when beginning the interview. These written materials should include:

- A description of A, the training device.
- A clear statement of the prediction target, T(A).
- A scenario for T(A). It will probably also include a brief description of the actual equipment system that the proposed training device is being designed to train for, as well as the device's configuration, hardware, and plan of instruction.
- A list of potential comparison cases, Bs. You need only brief information on each, since the SMEs should select only comparison cases with which they are familiar. When preparing a report, however, you may need to describe these comparison cases to others, so you should obtain some description.
- The checklist of high drivers. A brief explanation of each may be needed.
- Baseline data for T(B). If these will be estimated by the SME, there is nothing to gather in advance. If there are already some data, either operational data or research findings, you may want to track these down before the interview.
- A glossary of CBP terms and definitions.

It is important to prepare a data collection form prior to the interview. You may want formal sheets for entering the data and judgments (see Appendix B for a sample format), and an auxiliary notebook for recording comments. It is very easy to lose data if you are not prepared to record answers and changes in judgments. You will also want access later to important comments that the SME may have added during the interview.

The data form should include relevant background information on the SME. This will enable you to document his/her credentials as an "expert" in the area of interest. You also want to be able to locate the SME if follow-on questions are needed later. If the SME is military, you should ask the length of his/her current tour of duty and where his/her next posting is anticipated.

Before you use the CBP method for the first time, you should try a practice application. This will let you test your plans for conducting the interview and collecting the data. It will help you to think through all of the decisions concerning strategy that the interview process requires.

CONDUCTING THE INTERVIEW

There is a clear set of steps for the interview, leading to the objective of obtaining a prediction (see Figure 5). If there is to be more than one interview, the guide is important also for the sake of consistency from one to another. As with most face-to-face exchanges, however, it will be modified as it proceeds. The order in which the steps are taken may change, or some may be repeated as the SME comes to understand the process. Therefore the first requirement is an interview guide, prepared in advance, to assure that you do not omit any steps and that you can describe how you conduct the interview.

FIGURE 5

The CBP Interview

Step 1 Introduce task.

Step 2 Describe A.

Step 3 Determine B.

Step 4 Identify high drivers.

Step 5 Establish T(B).

Step 6 Judge global T(B) - T(A) differences.

Step 7 Specify T(B) adjustment and predict T(A).

Step 1 will be an introduction to the SME of the task and the CBP process. You may want to review briefly the guide to the interview that you have prepared. This is an opportunity to find out what you need to about the SME's background and also to answer any questions he or she may have. It should serve to put the SME at ease and to allow you some assessment of the level of understanding of the SME.

Step 2 has you describing the target training device A and the system for which it is to provide training. Explain the target measure T(A) you are trying to predict. Then describe the scenario or frame that you have developed for this measure. The SME may have some suggestions here, or objections to what you have devised. Be sure to note these comments. If you decide to revise T(A), be sure to specify on your data form just what you have done.

For Step 3, discuss with the SME the choice of comparison case B. You may settle on one from your list of possibilities or use a new one that the SME suggests. A short discussion should establish whether the choice is a suitable one, and the decision should be documented.

Step 4 requires you to identify T(B), the measure comparable to T(A) for comparison case B, and, most important, to present operational data for it. If there are reported field data on the effectiveness of device B, these will, of course, be used. Next best would be research data on B. (Steps 3 and 4 include tasks you will have addressed before the interview using other consultant experts.)

In Step 4 you will ask the SME to review your list of high drivers and amend it as necessary. The SME may add important causal factors of which you were unaware, or delete one or more which do not usefully pertain to B.

More often, though, no such data exist. In that case, Step 5 includes your asking the SME to estimate these data. Experience has shown that SMEs do not have difficulty in providing estimates. Give the SME the scenario you have developed for the mesure of T(A), and have T(B) data estimated for the same situation, as much as possible. Be sure to note whatever the SME wishes to say to describe how the estimate was made and what qualifications it may have.

You have now reached Step 6, the critical phase of the interview -- asking the SME to predict the relative effectiveness of A and B in terms of T. The SME will estimate whether T(A) is likely to be higher or (larger) for the target case A than T(B) is for the comparison case B — or whether T(A) is likely to be lower (or smaller) than T(B). This relative judgment is based on what the SME knows about B and what you have told him or her about A.

At this point, all you want to know is how the target measure T(A) and the comparison data T(B) stand in relation to each other. Later you will ask for an assessment of critical differences between the two cases to obtain a more nearly precise prediction. Step 6 is an overall judgment of the value of T(A) relative to T(B), that simply establishes two benchmarks for the final prediction: these benchmarks are the T(B) and the direction of change from it. This step is necessary because analytical judgments of high drivers can interfere with global judgments, so the latter are made first.

For Step 7, depending on the strategy you have chosen, the SME will refine and adjust T(B). No two applications of CBP are the same and no model can outline exact procedures. Users will have to be flexible and creative in tailoring each case to the particular need. With experience, you will learn to follow the line of reasoning your SMEs devise, and to tailor your interview to the specific situations. The outcome is T(B) adjusted to produce T(A).

We have written this section without examples, for the sake of focusing on the procedures. Appendix C details three case studies, examples of the application of CBP to three problems, using different strategies, and detailing the interview procedures and the method of adjusting T(B) to predict T(A). They illustrate how elements within CBP change to fit specific needs and take advantage of available resources.

CHAPTER VI: DOCUMENTING THE PROCESS

At this time, you have finished the main part of the task. You have completed the interview with the SME. You have generated a prediction for T(A). Now you must document the process.

The reason for documenting the prediction is so that others may understand how the prediction was generated. This can enable people to evaluate the prediction, or improve it, or update it as the design process moves on.

Your documentation should be concise. You need to describe the SME you interviewed (sometimes actual names would be used, but you need to be careful here if you have obtained information from an SME under an agreement of privacy). You want to name and perhaps briefly describe the comparison cases used, list the ways that the comparison case differed from the target case, and present the magnitude of these differences if you have obtained such information. Last, you would present the prediction of T(A).

It is very likely that you have recorded additional data on your collection form, information that does not fit specifically into the the prediction process. These may need analysis and presentation as well. For example:

> -- SMEs may have made comments that will be of interest and concern, often about design features of the training devices under consideration.
>
> -- SMEs may have made comments that will be useful in planning new training programs for the systems in question.

For most applications, 1 to 2 pages will be sufficient. If the documentation runs to more than two pages, begin with a brief summary or abstract outlining the salient points.

You may want to present the prediction with a range for T(A).

There are several ways to do this:

When an SME is estimating the magnitudes of the effects of the high drivers, you can ask for plus/minus values around each magnitude. At the end, you can add up the magnitudes to get the prediction, and you can add up the plus/minus values to get a prediction range.

For example, if there are three causal factors, the SME may say that the effect of the first is to add $10K dollars plus or minus $2K; the second may lower price by $50K plus or minus $5K; the third should lower it by $50K plus/minus $10K. The result is a prediction of a cost $90K lower than the comparison:

+$10K	+/- 2
- 50K	+/- 5
- 50K	+/- 10
-$90K	+/- 17K

If you have more than one comparison case, you can average them together, and use the range from the lowest to the highest estimate as the confidence range. If you use more than one SME, you can use the range of their predictions as the confidence range around a median. Other techniques may be used as well. None of these techniques meet the definition of statistical confidence limits, but they all reflect an ordinal scale of confidence. Users can refer to these ranges where there are a number of predictions, to see which predictions to rely on and which to be skeptical about.

APPLYING CBP TO COST PREDICTIONS

Some form of Comparison-Based Prediction is already the basis for most cost estimates that you make. You see what the same or similar items or programs cost and adjust that figure based on factors such as quantity, inflation, or materials.

If you have personnel inexperienced in cost estimates, then applying the CBP principles in this manual may be helpful in structuring that process for them. If you are practiced in estimating costs, this may provide you with some additional strategies.

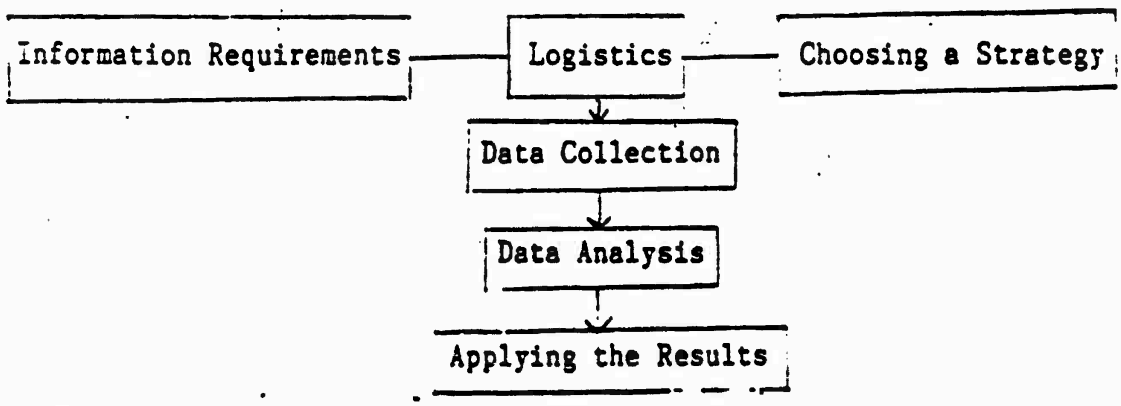
What CBP can bring to common cost estimating [illegible]

power of its elements, such as comparison cases and high drivers, and their identification and documentation. Strategies using multiple comparisons cases or multiple SMEs could add considerable strength to any cost prediction.

> For example, the HIP study applied CBP methods to the problem of predicting cost of a training device for an improved howitzer still in the very early design stages. Three different comparison cases were located yielding different cost comparisons. The analysis of causal factors, documented in the study, pointed clearly to the cost implications of different design options.

CHAPTER VII: MANAGING THE CBP PROCESS

As with any operation, the CBP process requires management effort. Different steps in the CBP process require coordination, approvals, scheduling, the assembling of relevant data, and the preparation of written materials. These needs of course will vary with the specific task and the design of the particular study you are undertaking.



LOGISTICS

Here are some logistics issues you should be prepared to manage:

— Contacts for the various information needs you have identified.

— Approvals for access to data, freeing time of SMEs, use of facilities, and other needs.

— Arrangements for meetings, interviews, surveys, equipment inspections, and other interfaces.

— Scheduling of the process, time estimates, follow-up sessions, briefings, and other blocks of activity. It takes at least 30-40 minutes to perform one CBP interview — up to an hour. It takes 5-10 minutes to explain the task and the reason for collecting the information, another 5-10 minutes for background information, and another 20 minutes to obtain the prediction, including estimates for the impact of the high drivers. When collecting data for several target cases, and several target variables, a data collection interview can run up to two hours. A rule of thumb is to allow 20 minutes for each specific comparison needed.

CHECKLIST OF INFORMATION REQUIREMENTS

__ What is the Actual Equipment Trainer (AET)? Usually, a training device is proposed to replace training with a real piece of equipment. You will want to prepare a brief description of this AET to show to SMEs who might not be familiar with it.

__ What is the target device? You will want to prepare a description of what you currently know about the proposed training device. This includes its configuration, hardware, and plan of instruction.

__ What are the potential comparison cases? You do not need to gather much information on each, since the SME should select only a comparison case that is familiar. However, in preparing a final report you may need to describe these comparison cases to others, so you should obtain some description.

__ What are the baseline data, the measure of T(B)? If these will be estimated by the SME, there is nothing to gather in advance. If there are already some data, either operational or research findings, you should track these down before the interview.

__ Who are the SMEs? You may need to collect some general information on the SMEs available before selecting the one(s) you use. Furthermore, you may need to justify your choice(s) later, to explain why you included a specific SME in the study.

__ What is on the checklist of high drivers? You will want to prepare a brief explanation of each high driver, so that you can show the SMEs how it differs from the others.

__ What is the target variable? Will you need any graphic aids to explain it to the SMEs?

__ Do you have a frame/scenario, for T(A)? The SMEs will need to look at this to understand the measure of training effectiveness T.

CONCLUSIONS

The emphasis of this Guide has been on explaining to you how to use the CBP method for making predictions about the cost and effectiveness of new training devices. CBP helps make predictions early enough in the design and development cycle that planners can make the right decisions while it is early enough to carry them out. CBP is a cost-effective way to make early decisions about changes in training device design and use, and about production and distribution schedules.

If a new training program is needed for a new weapons system, CBP can go back to the closest comparison case (usually the predecessor system) and assess the differences between the two so as to make design recommendations for the new program. This is an obvious use, and it is probably already widespread as an informal technique.

Another source of comparison is to identify training programs and training devices in other services and in industry that are comparable to the ones being considered, and to use the operational experience with these programs and devices to make recommendations about what should be included and excluded.

CBP formalizes these practices. It strengthens them by requiring the identification of high drivers and by documenting the process of adjusting data from the comparison case. It brings together the data and the SME for a sounder judgment.

In short, CBP can be used to take advantage of the operational experience that already exists. Rarely will a training device be entirely unique. If you can find out the lessons learned elsewhere, and systematically apply them to a new training system, then you will be able to benefit directly from past experiences.

REFERENCES

Hoffman, R.R. (1980). Meteaphor in science. In R.F. Honeck and R.R. Hoffman (Eds.) Cognitive psychology and figurative language. Hillsdale, NJ: Lawrence Erlbaum Associates.

Klein, G.A. (1980). Automated aids for the proficient decision maker. 1980 International Conference on Cybernetics and Society, October.

Klein, G.A. (1981). A perceptual/recognitional model of decision making. Paper presented at Summer Computer Simulation Conference, Washington. D.C., July.

Klein, G.A. (1982). The use of comparsion cases. IEEE 1982 Proceedings of the International Conference on Cybernetics and Society, 88-91.

Klein, G.A. (1984). Comparison-based prediction of program impact. Paper presented at summer meeting of Ohio Program Evaluators' Group, Columbus, OH.

Klein, G.A. (1984). Effects of time restriction on chess performance. Paper submitted for publication, April.

Klein, G.A. (1985). Paper in progress.

Klein, G.A., & Gordon, S.E. (1984). Using comparison-based methods for prediction and designing. Proceedings, Psychology in the DOD Symposium.

Klein, H.A., & Klein, G.A. (1981). Perceptual/cognitive analysis of proficient cardio-pulmonary resuscitation (CPR) performance. Paper presented at the Midwestern Psychological Association Meetings, Detroit, MI.

Klein, G.A., & Peio, K.J. (1982). The use of a prediction paradigm to study proficient performance. Paper prepared for Honeywell, Inc.

Klein, G.A., & Weitzenfeld, J. (1982). The use of analogues in comparability analysis. Applied Ergonomics, 13, 99-104.

Tetmeyer, D.C. (1976). Comparable item approach to establishing frequency of maintenance and maintenance tasks for a new aircraft. Wright-Patterson AFB, OH Aeronautical systems Division, Crew Station, Escape and Human Factors Branch.

Weitzenfeld, J. (1984). Valid reasoning by analogy: technological reasoning. Philosophy of Science, 51, 137-149.

Widenhouse, C., & Romans, W.W. (1977). A forecasting technique for operational reliability (MTBF) and maintenance (MMH/FH). Technical Report No. ASD-TR-77-28, Aeronautical systems Division, Air Force Systems Command, Wright-Patterson AFB, OH.

APPENDIX A

COMPARISON-BASED PREDICTION

Comparison-Based Prediction (CBP) is a methodology for making predictions when there are unknown parameters, missing data, or unclear objectives. Operationally, it is a way of structuring the judgments that experts make when they are called on to estimate unknown properties of a new situation. Formally, it is a system of reasoning by analogy, predicting to an unknown case by using what is known about a comparable case.

Reasoning by analogy is a natural process that contrasts with Bayesian and statistical decision models, which are hard to apply to most operational contexts. There is much current research into the importance of metaphor and analogy in creative thinking and decision making (Hoffman, 1980). Klein Associates has focused on the way experts use analogies to make inferences and decisions (Klein, 1985). We have studied Master chess players (Klein & Peio, 1982), experts in cardio-pulmonary resuscitation (Klein & Klein, 1981), and are currently working with firefighting chiefs.

A common example of this use of analogy comes from real estate. A realtor sets a price for a property, not by using a formal model and calculating all the variables, but by choosing a comparable sale and adjusting its price on the basis of differences between the two properties. Engineers have traditionally made use of analogies in prediction and design. They typically look for structural comparison. If their task is to predict how reliable a new piece of equipment is going to be, engineers use historical data for a basis of estimate.

A formalized version, called Comparability Analysis, is found in the Air Force (Tetmeyer, 1976). Developed in 1971, it is a way of explicitly using historical data to predict equipment reliability for the purposes of spare parts purchasing, manpower need projections, downtime forecasts, etc. Working directly from Air Force maintenance data, the engineer identifies a craft comparable to the one being planned. The next step is to derive an adjustment factor that reflects the differences between the comparison case and the new equipment. The third step is to present the rationale for the adjustment factor. Next is to collect the operational data for the comparison system, showing how reliable that equipment has been under operational conditions. The last step is to adjust these historical data to generate a prediction.

In the decade since Comparability Analysis was developed, it has been applied in the Air Force to a variety of new aircraft. Each involved a variety of subsystems, so that hundreds of these studies have been conducted. Widenhouse and Romans (1977) collected evaluation data contrasting predictions with observed data. When we analyzed these data (Klein and Gordon, 1984) we found that for mean-time between failures, the correlation between predicted and actual data was .76; for maintenance man-hours per flying-hours, the correlation was .84.

Klein Associates assessed the process of Comparability Analysis (Klein & Weitzenfeld, 1982) and presented an explanation of the logic underlying the use of comparison cases to derive predictions (Weitzenfeld, 1984; Weitzenfeld & Klein, 1982). We were interested in proving the method and in increasing its range of application beyond reliability and logistics. We studied three existing models of analogical reasoning, and found that none seemed to reflect the important aspects of Comparability Analysis: choosing an appropriate analogous situation; assessing the difference between it and the situation under study; and deriving an inference (prediction) by adjusting data obtained from the analog.

We have suggested a model of analogical reasoning that emphasizes the role of causal factors (Klein, 1982; Weitzenfeld, 1984). This model states that for Situation A there is a set of causal factors (x,y,z..) that will determine or influence T(A), the target characteristic of A to be estimated. Situation A could be a new aircraft duct system; causal factors x, y, and z could be the size of the aircraft, the material used, and a particular construction technique; and T(A) could be the reliability of the system as measured by Mean Time Between Failures.

In determining the target value, T(A), we usually cannot identify all of the causal factors involved, their effects and interactions. Instead, an analogous situation or comparison case (Situation B, another duct system) is identified which reflects the same determinants as the target case. That is, for aircraft B, the same causal factors (x, y, z..) determine a corresponding value, T(B), as a measure of system B.

Although the same causal factors affect both T(A) and T(B), it is unlikely that the values of the causal factors will be the same in both cases. In using T(B) as an estimate of T(A) we can take note of the differences in the values of each of the causal factors and make adjustments in our predictions to take these differences into account. Although checklists of causal factors can be provided, the method requires experts to use their experience in identifying the most important causal factors to use.

Comparison-Based Prediction is the methodology following from this model. The general CBP strategy (outlined in Table A-1) begins with the definition of the target variable, T, and the identification of major determining (causal) factors known to affect it. Next a selection of possible comparison cases is identified.

From these, Subject Matter Experts (SMEs) choose one case, based on the similarity of the effect of the causal factors between it and the target case. The comparison case value that is analogous to the target case value T(A) is specified as T(B). SMEs then make a rough estimate of the differences expected beween T(B) and T(A), most often only a judgment of whether T(A) will be greater or less than T(B).

Table A:1

The CBP Method

Setting up the Problem	-	Defining the Problem in concrete terms: T(A) = the Prediction Target
		Framing the questions: the Causal Factors and Prediction Scenario
Selecting the Resources	-	Choosing the Comparison Case and T(B) data
	-	Choosing the SME(s)-- Subject Matter Experts
Collecting and Analyzing the Data	-	Interviewing the SME(s) Analyzing Causal Factors to obtain T(A) from T(B)
Documenting the Process	-	Recording the Process
	-	Leaving an Audit Trail for others to follow

They then are guided through an examination of the effect of the expected differences in values of causal factors, until this effect can be quantified so as to produce an applicable adjustment factor figure. This factor is then applied to operational data for T(B), to yield a prediction for T(A). Analysis of the differences among factors produced by SMEs can produce a confidence range for the prediction. The process is documented to provide an audit trail, so that the basis for the prediction can be understood and the findings adjusted should changes be made in the target case.

The CBP technique relies on the use of Subject Matter Experts (SMEs) who are knowledgeable about the domain of interest, in order to select optimal comparison cases and identify the relevant causal factors. The CBP approach elicits SME judgments through the use of a carefully structured interview with a format reflecting the CBP process outlined in Table A-1. The approach is data driven since the SMEs are generating adjustment of operational data and giving their reasons for making these adjustments. There may be cases where no operational data are available. It is possible to proceed with a CBP approach by having the SMEs estimate the operational data, but this is not the ideal application of CBP method, and will reduce confidence in the outputs. However, this is often the state of affairs for the predictions where CBP is used, since this is usually the clearest case where there are no alternative prediction methods.

An important element in the CBP strategy that can increase our confidence in the prediction is the development of an audit trail. The audit trail consists of a detailed description of the causal factors considered by the SME, and the impact estimated for each. By having an explicit set of causal factors to consider in determining adjustments, the SME has a set of concepts to use in posing the differences between the target case and the comparison case(s). This facilitates communication among SMEs and helps to standardize the variables considered in the prediction process. In addition, if the prediction is found to be inaccurate once operational data are obtained for the target case, the audit trail provides an opportunity to go back and see which considerations (causal factors) were responsible for the misjudgment. This process is obviously not possible when only an unstructured expert opinion has been obtained.

It can be seen that CBP has several advantages over traditional prediction techniques. The CBP strategy is relatively easy and straight-forward, and can be used even when there are unknown parameters, missing data, or unclear objectives. In addition, it requires relative judgments from the SME (evaluating one situation in relation to another), which seem easier for them to make than absolute judgments. Perhaps the most important strength of the method is that it grounds the predictions in concrete experiences. Additionally, CBP creates an audit trail of the prediction process, which can later be used to evaluate and improve the prediction. Finally, it has high face validity in that it seems to be a structured form of a naturally occurring inference process, reasoning by analogy.

How does this method compare with the traditional use of expert judgments? CBP does use expert judgment, but in a structured, definable, and traceable manner. This allows for a clear and explicit basis to all judgments. The structure and definition are created by asking the Subject Matter Experts (SMEs) to identify relevant comparison cases and to make pair-wise comparisons and relative judgments. For example, CBP would direct the SMEs to select those training devices that they are familiar with that are most like the to-be-developed training device, in terms of major causal variables. By asking the SME whether the new training device is likely to be more or less complex (e.g., type of technology, or number of tasks to be trained) or different in availability (e.g., one-per-class vs. one-per-trainee), CBP structures the process of estimating effectiveness so that better judgments can be made. In contrast, asking Subject Matter Experts to estimate the effectiveness of a new training device, without the benefit of CBP methods, can result in judgments based in a wide variety of approaches and experience that may be more or less valid and comparable.

One brief validation study has been conducted to date. CBP was used to predict the outcome of an experiment on differences in effectiveness of functional and physical fidelity of training devices. Correlation of CBP predictions with test results was .90, accounting for 81% of the variance.

A major limitation of Comparison-Based Prediction is that it requires data about specific cases, not merely statistics about groups of cases. This data base is not always available but can usually be estimated satisfactorily by SMEs. This limitation can, however, also be viewed as an advantage: it combines 'real world' data with expert judgment, and this combination may produce more accurate estimates. The CBP methodology utilizes the skills and experiences of the SMEs for this purpose.

A prediction technique should be able to use most of the information available, such as causal relationships, and introduce few additional limitations or distortions. CBP fares well on these criteria. The following cautionary limitations of any prediction technique should be noted:

A. Requirement of expertise. Expertise is needed to describe the causal relationships, to identify comparison cases, and to plan and execute strategy in using comparisons. Inexpert application can make any technique ineffective.

B. Vulnerability to unknown effects. If a variable is not known to be influential, it cannot be taken into consideration. This is true of all predictive methods. CBP has the advantage of taking into consideration variables whose relevance is unclear, since such variables are already embedded in the comparison case data.

C. Incomplete data. In practical predictive situations resources are normally limited. Inadequate data and erroneous data will degrade any predictive process. One partial check for erroneous data is the use of redundancy. CBP has the potential for employing various combinations of analysis and correction to improve reliability. Moreover it calls attention to the simplifying assumptions made in the absence of complete data, and trade-offs between effort and validity should become apparent.

APPENDIX B
SAMPLE FORM 1

Menu: Possible Comparison Cases

Target Device	Comparison Cases	Relevant Features	Where In Use
	1a X X X	- - -	- - -
1. X X X	1b X X X	- - -	- - -
	1c X X X	- - -	- - -
	2a X X X	- - -	- - -
2. X X X	2b X X X	- - -	- - -
	2c X X X	- - -	- - -
	3a X X X	- - -	- - -
3. X X X	3b X X X	- - -	- - -
	3c X X X	- - -	- - -

Checklist of Causal Factors/High Drivers

Identified Causal Factors	High Drivers
Physical Fidelity	x
Task Complexity	-
Feedback Potential	x
Adjustable for New Tasks	-
Level of Trainees	-
*Reliability of Training Device	x

*Added by SME

APPENDIX B:
SAMPLE FORM 3

Interview Guide/Data Sheet

<u>Step 1</u> Introduce task.

<u>Step 2</u> Describe A.

<u>Step 3</u> Determine B.

<u>Step 4</u> Identify high drivers.

<u>Step 5</u> Establish T(B).

<u>Step 6</u> Judge global T(B) - T(A) differences.

<u>Step 7</u> Specify T(B) adjustment and predict T(A).

STEP 1

1. DATE____________ 2. SME #____________

3. NAME______________________________________

4. TELEPHONE__________________ 5. HOW REACH LATER__________________

****[Introduce Self]

6. JOB__________________________

7. Length of Experience: in Job__________________________

with TD (B)_____________________

STEP 2

****[Introduce Task: a) Need to predict T(A) for new TD
b) Method = adjust analogous case data]

8. Define scenario and T(A): # of hits in final test round in class____

Is this a reasonable measure? How must it be re-defined or refined?

____

____

STEP 3

9. Show list of comparison cases. Agree?_____Change Choice?_________

STEP 4

10. Describe A. Present <u>Checklist of Causal Factors/High Drivers</u> and record answers to 13 and 14 here.

<u>High Drivers</u>	<u>T(A): + or -?</u>	<u>Absolute or %</u>
Physical Fidelity	____	______
Feedback Potential	____	______
Task Complexity	____	______
Adjustable for New Tasks	____	______
Level of Trainees	____	______
Other:________________	____	______

11. Ask if there are additional causal factors to be considered (add above).

STEP 5

12. Review T: What is the value of T(B) in your experience?

____

STEP 6

13. For each, ask SME to judge direction of difference (+ or -) for T(A). (Record in 10 above.)

STEP 7

14. Ask if can estimate <u>degree</u> of difference for T(A) on each causal factor. (Record in 10 above.)

____

15. Adjust T(B) to T(A) on basis of above.___________________________

16. This is your prediction -____

Are you comfortable? Any Changes?_____________Any Provisos?______

APPENDIX B:

SAMPLE FORM 4

Documentation

The documentation of a CBP application usually takes the form of a brief written report. The checklists and other interview forms are attached and referenced. If multiple SMEs and/or comparison cases have been used, the data are presented in tables.

The briefest documentation could take the following form:

1. Problem:

2. Prediction Obtained: T(A):
   Discussion of meaning of T(A)

   SMEs comments on TD design, training implications, etc.

3. Summary of CBP Process:
   SME background data

   Menu of comparison cases showing those considered and those used

   Checklist of causal factors showing those use,not used, and added by SMEs

   Interview form

   Table of SME judgments on effect of each causal factor

   Table or figure showing calculations of T(A) and confidence ranges.

   Listing of SME comments that have relevance for TD design or training program.

APPENDIX C: THREE CASE STUDIES

[NOTE: These three case studies are excerpted from research undertaken for the Army Research Institute, assessing the feasibility of the CBP methodology for problems of predicting training device cost and effectiveness. You are encouraged to read the full reports of these studies for a more detailed presentation of the development of each CBP element and the application of the method.]

CASE I: The High Driver Strategy

In this instance, an automotive maintenance instructor was asked to predict the time that might be saved in his course if a training device, the Army Maintenance Trainer Evaluation Study (AMTESS), were available for class use. The following elements of CBP strategy were already decided upon in preparation:

- The target value T(A) was determined to be minutes saved in Course 63W10 if the AMTESS was used.

- The comparison case was chosen to be the course as taught currently, which used the actual automotive equipment (AET) for instruction.

- Five causal factors were identified.

  - training potential of the device;
  - its availability;
  - a utilization factor that covered issues such as the efficiency of the device for the training tasks;
  - the degree to which the device was judged to raise students' motivation for learning; and
  - time needed to learn the training device operation.

The SME was asked to consider the target variable T(A): minutes saved in training for trouble-shooting of the engine starting system. He first estimated the number of minutes that were now spent in teaching this task to criterion, using only the actual equipment for instruction and practice. This judgment produced the comparison case measure T(B).

He was then asked to consider the effect on T(B) that using the training device would have. The SME was not asked to make this judgment in one step. He was asked first to make a global judgment: if using the training device would increase, or decrease, or leave unchanged the number of minutes needed to train this task to criterion? Then he was asked to consider, one at a time, the five causal factors identified as driving considerations in this judgment. He then determined an adjustment to the total time based on each factor.

When asked to consider the effect that the first causal factor--training potential of the training device--would have on training time for this task, (training for engine-starting trouble-shooting) the SME was led through the method as follows:

1) How many minutes are used for training this task as the course is now constituted?

2) Consider using the AMTESS training device in this course instead of the actual vehicle. As you judge the training potential of the AMTESS training device compared to that of the vehicle itself, would using the AMTESS training device produce a savings of time in the course? Or would it require additional time to teach the task? Or would it make no difference in the time needed to teach this task to criterion? Express your answer in actual minutes saved or added.

3) If using only the AMTESS training device and not the vehicle itself means you cannot train to criterion, you can consider using the AMTESS training device with some supplementary training on the actual vehicle. How much time would that add to your estimate?

4) Now consider the factor of availability. Knowing what you do about the way the AMTESS training device availability compares to the actual vehicle, would using the AMTESS training device save or lose time in the training for this task?

The SME was led through this procedure for each of the remaining three causal factors: utilization, motivation, and training device learning time. With each, T(B) was adjusted for the first teaching task. The procedure was repeated for other tasks within the course, such as training the trouble-shooting for an oil pump failure. His answers on each factor were summed, and added to T(B) to yield T(A), a prediction of the number of minutes it would take to teach this course if the AMTESS training device were used. The SME's estimate for each causal factor was a measure of time saved (or added) in the course.

Comparison of T(B) with T(A) showed a net savings on this task of 17% of training time if the AMTESS training device was used, supplemented by actual equipment. The major adjustment that led to the savings was on the factor of motivation. The SME thought that the training device was designed so that students would participate in the class to a degree that would save considerable training time. The table below represents an "audit trail," an element of the CBP method which assures that the prediction can be examined.

HIGH DRIVER STRATEGY: Results

Course Time Savings Prediction: AMTESS training device Prototype, Course 63W10
[Estimate of T(B)*: current total course time = 450 minutes]

CAUSAL FACTOR	MINUTES SAVED
Training Potential	-30
Availability	30
Utilization	0
Motivation	100
Training Device Learning Time	-25
OVERALL SAVINGS	75/450 = 17%

*The Actual Equipment is the comparison device, used in 63W10

Other forms from the AMTESS study are attached (Figures C:1 - C:4), to show you the variety the CBP method produces. Remember, CBP is an approach, with basic elements that you are to apply to your own situation. Start with these elements and design your own forms to hold the information you need.

Comparative Matrix of AMTESS Design Characteristics

		GRUMMAN	SEVILLE
1.	Principle of Simulation	Simulated, generic	Simulated, actual(models)
2.	General configuration	Independent student control station	Satellite, instructor mediated
3.	Instructor station	Same as student station	Separate instructor station
4.	Student stations	Single-student	Single-Student
5.	Graphic display	Color CRT video disc 17" tube	Slide Projector Screen & CRT
6.	Alphanumeric	BW CRT 12"	CRT BW
7.	Input device	Touch panel (finger)	Function keys on response panel
8.	Keyboard	Required for instructor only	Required of instructor only
9.	Test equipment	Actual measurements simulated T.E.	Actual measurements simulated T.E.
10.	3D Hardware	Family generic, context* preserved	Family, models, context* preserved
11.	Operating System	Pascal, fixed RAM, convertible to ADA	RT-11 DEC
12.	Program storage	Floppy disc	Winchester Disc w/floppy backup
13.	Program language	Pascal (convertible to ADA)	ARIC (Burtek proprietary)
14.	Authoring system	Instructor-easily used	Instructor, medium difficulty
15.	Portability	High	Medium
16.	Interfaces	2D-3D, selectable to 110K baud printer	Printer RS-232 9600 baud
17.	Modularity	Good	Medium
18.	Diagnostics	To individual module	Test program, diagnostics for computer
19.	Repair procedure	Substitution of modules	None specified-from T.M.
20.	Motion	Full-plus freeze	None
21.	Audio	Full	None on radar, engine sounds on diesel
22.	Fault insertion	Simulated, program mediated	Simulated, instructor mediated
23.	Motor skills training	Via 3D simulation hardware	Model of actual equipment
24.	Actual equipment	Not required but usable	Not required not usable

*Context is that of a diesel powerplant

Figure
C:2

Course Tasks Used in Predicting Training Effectiveness

Grumman

63D30 and 63H30

Task 1: Perform tests on electrical system (continuity test with STE/ICE, resistance test with STE/ICE, DC voltage test with STE/ICE, AC voltage test with STE/ICE

Task 2: Troubleshoot electrical system (starting system, generating system, battery power system)

Seville/Burtek

63W10

Task 1: Troubleshoot engine starting system

Task 2: Troubleshoot oil pump failure (organizational and direct support)
- Perform organizational troubleshooting
- Perform direct support troubleshooting
- R/R oil pump filter and oil pump

63B30

Task 1: Troubleshoot oil pump failure (organizational T/S only)

Task 2: Adjust alternator drive belt

Task 3: R/R starter motor

Task 4: Inspect electrical system

Copy available to DTIC does not permit fully legible reproduction

Copy available to DTIC does not permit fully legible reproduction

Figure
C:3

Summary of SME Background

Maintenance Course Instructors	Location	Area of Expertise
SME 1	APG	Course 63D30
SME 2	APG	Course 63H30
SME 3	Edgewood Arsenal	Course 63W10
SME 4	Edgewood Arsenal	Course 63B30
SME 5	Edgewood Arsenal	Course 63B30
Air Force 6883 AIS		
SME 6	Air Force Human Resources Laboratory	Evaluation of Honeywell 3-D and Burtek 2-D simulators
SME 7	Honeywell	Design Engineer
SME 8	Honeywell	Research Psychologist
Prototype TD Evaluators		
SME 9	APG	Evaluated TD for automotive tasks
SME 10	Ft. Bliss	Evaluated TD for missile tasks

Figure
C:4

October 3, 1983

CBP Forms, AMTESS

Name ______________________________

Office Symbol and Telephone ______________________________

Date ____________________ Course Taught ____________________

Students ______________ Course Length ______________ A_t: ______________

AMTESS device: ______________________ Task: ______________________

Comparison device(s): ______________________ B_t: ______________

Target Configuration: ______________________________

Relative Impact: Target Configuration vs. Existing Configuration:

__________ Better __________ Same __________ Worse

Causal Checklist:

	Current Configuration	Target Configuration	AMTESS Hours
Training Potential (Procedures, Perceptual-Motor, Decision making, Task Integration, Other			
Utilization (Integration into POI, Ease of Operation, Set-Up time, Performance Evaluation, Instructor Aids, Ease of Modification)			
Availability (Reliability, Supportability, Repairability)			
Motivation			
Other			
Total			

Figure
C-5

SME Judgment of
Potential Impact of Causal Factors
on Training Time

Causal Factors	Training Time* (1) Student learns to use TD	(2) Train task on TD	(3) Supplemental Training on AET
(A) Original Set (Prepared before interview):			
Training Potential		+	+
Utilization	+	+	+
Availability	+	+	
Motivation	+	+	+
(B) Alternative Detailed Set (Developed in interview):			
Training Potential			
TD provides accurate feedback		+	
TD teaches use of the Technical Manual		+	+
TD shows where components are on the AET		+	+
Utilization			
Ease of programming malfunctions		+	
TD allows skipping of steps		+	
Computer response time	+	+	
Availability			
Reliability	+	+	
Ease of maintenance		+	+
Time required for maintenance tasks		+	+
General			
Difficulty of learning to use TD	+		
Information provided by manufacturer	+	+	
Use of videodisc		+	
Requires reading skills	+	+	

*Baseline is current course time. + indicates estimated addition to this time because of TD use.

(1) Time needed to learn to use TD itself.
(2) Time needed to train task on TD.
(3) Time needed for supplemental training on Actual Equipment.

CASE II: Multiple Comparison Strategy

In this example, a tank gunnery instructor (SME) was asked to predict the effectiveness of a tank gunnery training device Videodisc Gunnery Simulator (VIGS). The scenario provided him was to assume that the men were to practice on the VIGS during the six months between their school gunnery training and their first unit field exercises. The following elements of the CBP strategy were determined:

— The target value T(A) was defined as number of first-round hits on Gunnery Tables VI and VII.

-- Multiple comparison cases were chosen from among training devices that are used in gunnery training classes.

-- Multiple causal factors were identified in advance but the SME was to choose his own basis for comparison during the prediction process itself.

The SME was asked to consider the number of first-round hits the gunner would attain at the end of training. He then judged the number of first-round hits he would attain six months later with no intervening training or practice. This latter judgment would be a first, baseline T(B). Then he also estimated the performance of the best gunner imaginable six months after school training, to set an upper prediction limit.

The SME then was asked to consider, one at a time, other training device's, and to estimate the number of first-round hits the gunner might achieve six months after training if that training device had been used for practice for a specified period in the six month interval. These judgments yielded additional T(B)s (see illustration).

With these figures for comparison, the SME then predicted the number of first-round hits likely if the gunner trained on the VIGS device for the specified period in the six month interval. This figure was T(A). The differences between T(A) and the T(B)s was a measure of the relative effectiveness of the VIGS device (80%-72%, for example, for VIGS relative to "no practice").

One effect of the use of different comparison cases was to structure the prediction space that the SME used. By setting upper and lower limits, the total space was bounded. By positioning the different comparison training devices within that space, the locus of the target training device was more clearly defined.

In this instance, the SME was able to predict training device effectiveness under two test conditions (moving and stationary platforms) and for two tank environments (M60/A1 and M60/A3). When the predictions of several SMEs were compared, they converged closely on the same T(A).

MULTIPLE COMPARISON STRATEGY: Results

Gunnery Training Effectiveness Prediction, VIGS Training Device

CEILING: 92%
(Estimated,
best gunner,
no practice)

TELFARE: 85%
(estimated)

PREDICTION *
RANGE

MK60: 80%
** T(B) **

BREWSTER: 76%
(estimated)

BASELINE: 72%
(estimated,
average crew,
no practice)

* PREDICTION expressed as percentage of first-round hits attained six months after school training, at start of first unit field exercise

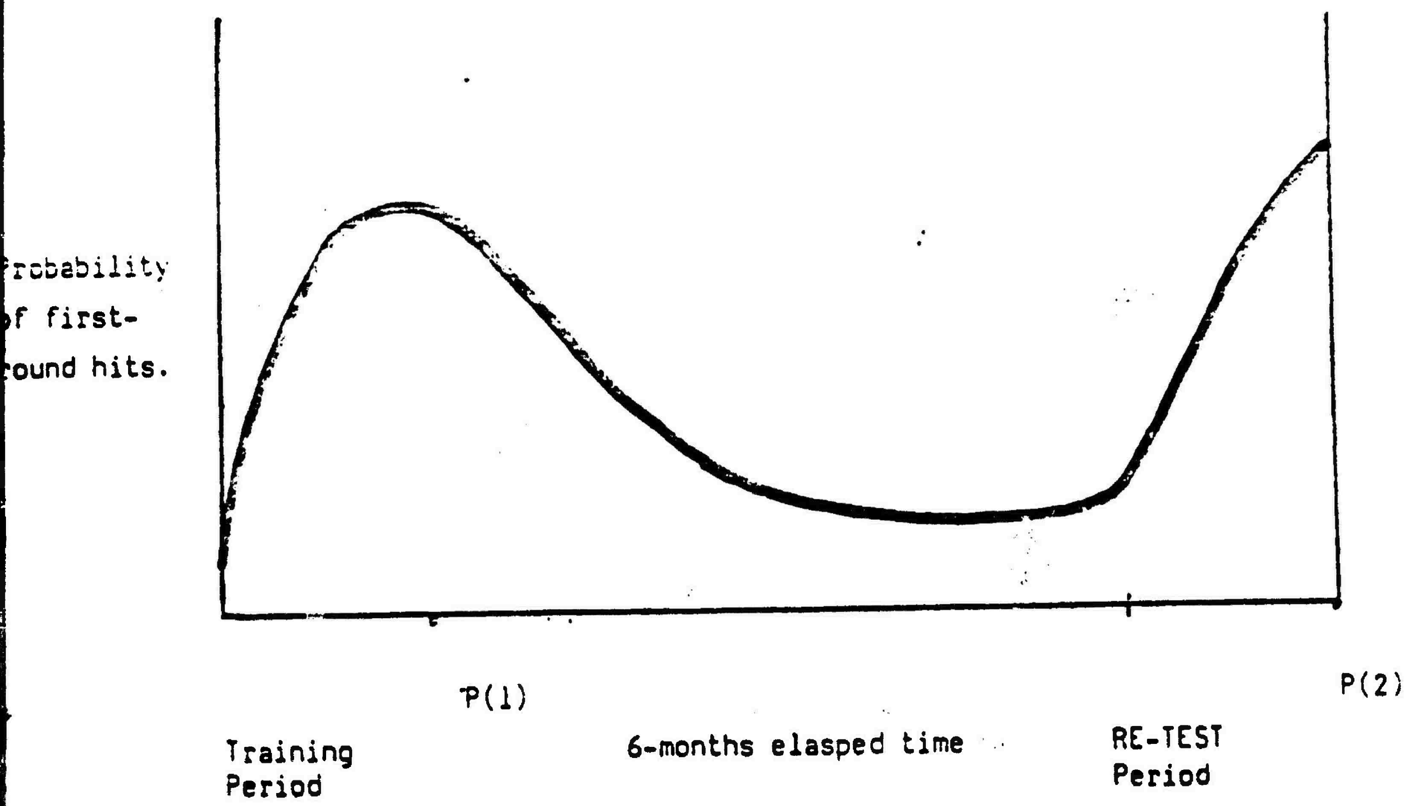
---

Other forms relative to the VIGS study are appended (Figures C:5 - C:9 Again, they reflect the application of CBP to this particular problem. They illustrate the flexibility of the method, and the way its elements can be adapted to different needs.

Figure

C:6

The Prediction Scenario: End-of-training Proficiency and Sustainment Proficiency



P(1) Estimate of % first-round hits when proficient at end of initial training

P(2) Estimate of % first-round hits on Table VI and VII after 6-month interval

Figure
C:7

Descriptions of Training Devices Used as Comparison Cases

1. The Telfare TD (TEL)
This device is tank-mounted, and includes a bracket which permits the use of the M2 .50 cal machine gun (in single shot mode) in lieu of the main gun. It can be fired against full-size targets with a curtailed range or against half-size targets, half range. All four tank crew members can practice coordination for both stationary and moving platform gunnery. The range finder can be used, including LRF, but there is no range correction for the half range mode. The device can be used with thermal sighting as well as the standard optical sighting. Laying the sight on target, tracking, firing, and fire adjustment can be practiced.

2. The Brewster TD (BRW)
This device is also tank mounted, and includes brackets permitting attachment of small caliber weapons (either .22 cal or 5.56 mm.). It is used only with stationary platform gunnery tasks. In the .22 cal version, a minified range (1/30 or 1/60 scale) is used with very small targets on a sand table (stationary and moving targets). Sight laying, tracking, firing, and fire adjustment can be practiced as well as overall crew coordination.

3. The Tank Gunnery and Missile Tracking System TD (TGMTS)
This device is mounted on a tank, but used inside a large indoor training facility. It includes a rear-projection screen depicting the battlefield, and a laser device mounted on the tank turret to simulate the main gun firing. It is used to train both TC and gunner to gain coordination and efficiency. Feedback can be provided to the gunner in the form of "frozen" action shots to show the exact point of hit. Tasks trained include stationary and moving targets, but only from a stationary platform.

4. Actual Equipment Trainer (AET)
As a final comparison case, some SMEs were asked to generate predictions based on comparison with the use of an actual tank with the firing of real main gun ammunition. In some ways this can be seen as the ideal training device. However, there are certain drawbacks to using the AET for training, such as cost, range availability, a lack of feedback, etc.

Figure
C.:8

SME Background Information

SME Number	Rank	Task Area	Previous Experience	Length of Experience in Training Others	TD Familiarity
1	Civilian (Ph.D)	ARI Field Unit	-	-	BRW TEL TGMTS
2	SFC	gunnery instructor	TC; 4 years Platoon Sgt.	5 months	BRW TEL
3	Major	DOTD*	-	3½ years	MK60** BRW TEL TGMTS
4	Sgt.	gunnery instructor	TC Platoon Sgt.	6 years	MK60** BRW TEL TGMTS
5	Sgt.	Weapons Dept.	TC	8 months	BRW TEL TGMTS
6	Sgt.	Weapons Dept.	TC	1 year	BRW TEL TGMTS
7	Sgt.	Weapons Dept.	TC Platoon Sgt.	6 years	BRW TEL
8	Sgt.	Weapons Dept.	TC Platoon Sgt. 6 years: Master Gunner	5 years	BRW TEL TGMTS
9	Sgt.	Weapons Dept.	TC	1 year	BRW TEL TGMTS
10	SFC	DOTD	TC Platoon Sgt.	3 years	MK60** BRW TEL TGMTS

*Directorate of Training and Doctrine

**A1 only: no SME had experience with MK60/A3

Figure C:9

Final Checklist of Causal Factors

Realism
- Visual
- Audio
- Feel
- Tank
- Targets
- Ballistics
- Variety of Scenarios

Utilization
- General
- Set-up ease
- Frequency

Training Value
- Instructional Aids
- Feedback
- Familiarity/Confidence
- Relaxation Factor
- Motivation

Functional Fidelity
- Gunner training
- Make procedures automatic
- Range of tasks
- Lead training
- Malfunctions

Crew Coordination
- TC ranging practice
- TC - gunner interaction
- Loader - gunner interaction

CASE III: Global Prediction Strategy for Acquisition Cost

This study applied the CBP methodology very early in the design process of the Howitzer Improvement Program (HIP). The new system itself was not yet firmly described, so the new training requirements were unclear.

The prediction variable, T, chosen was

- o Acquisition cost of a HIP training device.

Given the early stage of the design process, we were not sure how CBP would work. We found that it worked easily, because at this early stage, proposed devices were being identified on the basis of analogy to specific existing devices!

The procedure for proposing a device was largely that of identifying trainers for other systems and evaluating their utility for HIP.

Since tank gunnery training was going to be conducted in a new Conduct of Fire Trainer (COFT), one proposed HIP device was analogous to COFT. The question for us was: what would it cost? To answer this, we obviously used the COFT as the comparison. There were two versions: ICOFT (an institutional trainer) and UCOFT (a unit trainer). Since HIP would need an institutional training device, we selected ICOFT.

The SME chosen was in the Armor Division at PM TRADE. He had been involved in the development of ICOFT and the evaluation of UCOFT, and had 13 years Army experience.

He broke the definition of T into separate components for building, hardware, and RDT&E (research, development, training, and evaluation) which includes the development of instructional software. Next, he listed the high drivers most likely to affect the differences between the cases, for each of these components.

He judged that the factors likely to have most impact on building costs were heating, air conditioning, computer support equipment, and power distribution. For hardware costs he considered fidelity of the simulation, mechanical requirements for specific new tasks, the computer system size, the visual system, and the number of stations needed because of different student/teacher ratios that could be sustained. RDT&E costs would differ between the two system: much of the development of the old system would carry over for the new and did not need duplication; there would not be competitive procurement, reducing cost; and simpler software would be needed.

The SME was asked to make global adjustments to UCOFT costs — for which he supplied figures — in order to predict costs for the HIP device. Since he could judge the number of stations that would be required, he was asked to estimate the cost of a single station and then multiply by the number required to reach a total cost. The results are as follows:

Factor	COMPARISON CASE ICOFT	HIP PREDICTION
Building	$ 4,750,000	$ 2,500,000*
Hardware	3,600,000	8,000,000**
RDT&E	5,300,000	8,000,000***
	$13,650,000	$18,500,000

* Primary reason for difference: 16 student stations for proposed HIP device vs. 36 for ICOFT. This impacts air conditioning, heating, computer support equipment, power distribution.

** The HIP recoil mechanism will be very costly. On the other hand, the visual requirements for HIP will be less, as will the computer requirements, since HIP doesn't need to compute round trajectory in real time. Another factor favoring HIP is the ratio of 1 instructor station/4 student sations; it is 1:1 for ICOFT.

*** The ICOFT benefitted from $33M spent on RDT&E for UCOFT. The total COFT expenditure was $38.3M. The total for HIP is therefore much less than the total for COFT. The reasons are: the COFT project included an expensive competition between GE and GD, the howitzer problems are simpler -- no lead or trajectory concerns, fewer dynamics. The howitzer expense for turret recoil should be greater.

This process produced a clear audit trail, including documentation of the assumptions the SME made in his prediction, the comparison case on which it was based, what types of adjustments were made, and the factors that were considered in making them. This estimate can thus be revised as more factors become known. It can also be compared to estimates made by other means.